

Can research benefits be proved?

A US Congressman is asking whether support for basic research can be taken for granted. The research community, which is inclined to react impatiently, should instead give a moderate response to a fair question.

Mr George Brown, the chairman of the Committee on Science, Space and Technology in the US House of Representatives, has raised an inflammatory question (see page 175) at the worst possible time — when the US federal budget for the financial year beginning on 1 October has not yet been settled. The question is whether the federal government should continue to support basic research in the absence of proof that research yields commensurate social benefits. Researchers have inevitably responded hotly to the suggestion that the US government no longer owes basic research a comfortable living. This time, it seems, will be no exception; the scientific knives were out even before the report was released. But Brown deserves a more careful hearing, if only because his position may help him to turn his suggestions into public policy.

For the US research enterprise, the 1980s were remarkable. Successive administrations not conspicuous for farsightedness protected basic research from budgetary constraint quite astonishingly. The annual budget of the National Institutes of Health (NIH) is now almost \$9 billion, partly because of the tendency of Congress to outbid the administration in generosity. The National Science Foundation (NSF), only a quarter as well off, has also grown year on year, even if it has not doubled in size as it had planned. Meanwhile, the Department of Energy (DOE), the National Aeronautics and Space Administration (NASA) and the Department of Defense (DOD) remain significant sponsors of research, some of it basic research outside their own laboratories. Who can be surprised that the 1980s also saw US researchers reinforce their dominance in the most active fields of basic research, from astrophysics to zoology? How, against that background, dare Brown raise such a sour question?

The truth is that there is nothing wrong with asking the question but some ways of answering it may be disastrous. Britain's experience in the 1980s is proof of that. The then-new Thatcher government asked Brown's question and gave its own answer — that publicly supported research must be reorientated to yield more immediate economic benefits. That is the spirit in which it is this week recommended (see page 179) that Britain's already truncated research in thermonuclear fusion should be converted into industrially orientated applied research.

So researchers are demoralized because opportunities to break new ground have vanished. More recently, young people have shied away from university courses in science and technology. It is to be hoped that Brown will not encourage the British answer to the question he has asked.

What other answer should Brown be given? First, he should be reminded that his is not the only demand for the evaluation of the fruits of research. NSF, too, is warming to research tied to economic needs. Brown's proposal is also echoed by the promise in the strategic plan on which NIH is still labouring: that future programmes will be judged by their prospects of yielding social benefits, better health care specifically. NIH's dependants in the research community are still wary of this undertaking — and rightly so — while the proposed mechanism of assessment is not in place. Yet does it not make good sense that the draft strategic plan foresees, for example, that the Human Genome Project will be followed by a presumably even more expensive programme to investigate the cellular mechanisms of normal and aberrant genes? Even those whose interest is exclusively a better understanding of how cells function will not be inconvenienced by such a policy.

Brown at least seems to understand that the social and economic benefits of basic research do not rest solely in the published work that fills the scientific journals. The cultural value of lively research is not easily quantified but is not zero. More tangibly, education through basic research has become the acknowledged means of providing the force of skilled young people on whom the success of a modern economy depends. Who will put that in hazard while glooming about the future competitiveness of the US economy? And what of the way in which the research enterprise has become one of the most prolific stimulators of small business in biotechnology and electronics? Brown is right to ask the Office of Technology Assessment to put a figure on the value of that process.

Budget makers (and budget cutters) should be more aware of these side effects of basic research than is their custom. They should also acknowledge that the sheer quantity of money is often less important than its allocation and the way it is spent. That is as true for the US research enterprise as a whole as for the organizations



and the laboratories of which it consists. At what might be called the level of macroallocation, Brown will have noticed that Dr Bernadine Healy, director of NIH, has been complaining at NASA's commitment to the space station. She might as well have complained at DoE's Superconducting Super Collider, an innovative construction project with basic research attachments, which also steals funds that would be more productive in basic research. That does not imply that basic research should take precedence over all other spending on technical developments. It is rather that Brown and his colleagues should decide whether basic research incidental to the interests of operational departments of government provides the value for money they now look for from the grant-making agencies. (Amusingly, they would find that the Department of Commerce, through the National Institute of Standards and Technology, has been the engine of innovation in modern spectroscopy.)

At microallocation level, there are other questions to ask. Why is it now the general complaint of established researchers that their time is spent preparing applications for research grants? The doctrine that accountability for public money requires frequent reapplication is old-fashioned and bars the federal grant-making agencies from offering support of the kind the Howard Hughes Medical Institute has shown to be fruitful. And how correct is the common belief that the most successful applications are the safest, neither the most imaginative nor the most daring? The peer-review process, with all its virtues, tends to justify itself. And cannot the perpetual haggle about indirect costs be made at once more seemly and more equitable? NSF's experiment with grants in mathematics (see *Nature* **359**, 94; 1992) may help even if its main purpose is different. And how is the value of extramural and intramural research to be compared? If the education of students in research counts for anything, extramural spending should have the edge.

Brown is right to assert that times are hard, and that assumptions about the funding of research must be reexamined. Many of the questions he poses are good ones, and the fact that scientists find them threatening is not proof that they should not be asked. But at a moment when the political winds — at least at NSF and NIH — appear to have actually shifted his way, it is a shame that Brown has not started down a more reasoned and realistic path. Asking research to address social and economic concerns is fair enough; asking it to provide "greater opportunities for self-realization" is not. Social science may indeed have something to contribute to the process of assessing research, but it should not dictate its ends. Although US science has repaid its country well for 45 years of investment, the time may indeed have come to rethink the contract under which it has been supported. Pity then, that Brown's vague polemic risks inflaming scientists rather than opening their minds. □

Trade and environment

Alarms about the risk of food contaminants may become a powerful non-tariff restraint on trade.

WHOLESOME food and drink is everybody's wish, but what does wholesome mean? As things have turned out, most people are told by their governments what may be safely eaten or drunk. And most governments now have formal regulations defining the amounts allowed of accidental adulterants, from pesticides to radionuclides, that may be sold commercially or imported from elsewhere. The good sense of such arrangements is self-evident: government inspectors at ports are the most effective means of keeping off the world's markets animal carcasses infected with bacteria. But may not the same government inspectors restrain trade in agricultural produce that could not harm a soul, to the detriment of consumers and producers alike? That was a recurring theme at a conference in the Netherlands last week, organized by the Dutch government and the International Policy Council on Agriculture and Trade.

The issue is important because free trade in agricultural produce is the most effective way in which developing countries can work their way out of poverty. To be sure, we are a long way from the point at which that will happen. The rich countries of the world, which are ready enough to help make agriculture more efficient in developing countries, then meanly keep out the produce of the newly efficient farmers by means of tariffs and quotas. But that cannot last much longer. If Europe and the United States reach the accommodation on agriculture required for the current General Agreement on Tariffs and Trade (GATT), still in suspense, the next round of negotiations must include a deal on developing countries' farm produce.

The obvious danger is that, by then, the potential importers of foodstuffs from developing countries will have so protected themselves by finicky regulations that the imports will be deemed illegal and will be forbidden. Draconian standards on the allowable content of pesticides in foodstuffs, or insistence that imported fruit should have been grown from plants certified free from virus disease, would often be more effective than tariffs at keeping out unwelcome foodstuffs. In short, there is ample scope for unholy alliances between farm lobbies eager to protect their domestic markets and national environmental lobbies advocating the health risks associated with this or that contaminant. That is a powerful combination. The best defence against the distortion of trade that would result is an international understanding of the risks to human, animal and plant health of food contaminants. GATT, usually concerned just with numbers, should get ahead with reaching such an understanding, its other preoccupations notwithstanding. □

Brown starts debate with his report asking science to improve economy

Washington. When US Representative George Brown (Democrat, California) decided to write a report asking researchers to do more towards solving the nation's economic ills, he was looking for an argument. Thanks to an unauthorized draft shown to scientists attending a summer retreat last month, he got one sooner than he expected, setting the stage for lively hearings before his House science committee.

Science policy tends to be dry stuff, and congressional reports are plentiful. But Brown is counting on more than the usual rhetoric at a series of science policy hearings, and a provocative report seems just the thing to set the right tone. In a report* officially released this week, Brown and his staff call for a "rethink" of science priorities with a new emphasis on funding research that promises to contribute to "national priorities" such as economic competitiveness.

The idea is not new; in the 1970s, the US National Science Foundation (NSF) operated a programme called Research Applied to National Needs. Nor is it, by itself, much of a threat to the status quo. Although Brown is an influential figure in Congress, he does not have the authority to change significantly the way US agencies spend billions of dollars on science.

But Brown's words, coming on the heels of new initiatives at NSF and the National Institutes of Health intended to orientate research priorities towards national economic and social needs (see *Nature* 358, 355; 1992 and 358, 661; 1992), may become a lightning rod for protests. US researchers have rarely before been asked to defend their work on anything but scientific grounds.

"It's down to the last blank check", Brown wrote last week in the *Los Angeles Times* in an opinion piece timed to accompany the release of the report. "We've paid for 45 years of discovery; let's start requiring its application to the critical problems in the civilian sector." Brown wants to measure the value of research "not by number of publications or citations or patents, but by progress towards specific [national] goals", and some of his examples could further raise scientific eyebrows: "greater opportunity for self-realization", "less dependence on material goods as a gauge of wealth or success" and "less armed conflict".

Brown takes a somewhat more sober tone in his committee report, although his conclusion is the same. He says the government should define the national goals towards which research should be expected to contribute. Science agencies should com-

mission independent "performance assessments" to determine what research contributes most to the national goals. And the White House's Office of Science and Technology Policy (OSTP) should become the "command center" for implementing and evaluating all research policy decisions.

Those who read a leaked draft of the

"We've paid for 45 years of discovery; let's start requiring its application to the critical problems in the civilian sector."

US Rep. George Brown

"He blames science for what are really the failings of the political process."

David Baltimore

report at a meeting last month (an interdisciplinary conference held on 15–21 August at the Keystone Center in Keystone, Colorado) were incensed by its tone. The document "blames science for what are really the failings of the political process", says Nobel prizewinning biologist David Baltimore. "It misses the whole point about science being forward-looking, it is vague about its goals and the actions it proposes are pure bureaucracy." Many of the solutions Brown proposes as alternatives — such as block grants,

grants based on past performance and funding decisions by 'smart managers' — have been tried in Europe with limited success, he says.

Social scientists have also had little success in trying to calculate the impact of particular research projects. That aspect of Brown's proposal, says Roland Schmitt, the president of Rensselaer (New York) Polytechnic Institute and a member of the National Science Board, which oversees the NSF, is "going to require a lot of discussion".

Much of the outcry appears to be based on the source of the report; Brown has been a strong supporter of science in the US Congress and has often provided a counterweight to attacks from US Representative John Dingell (Democrat, Michigan) and others. The fact that Brown is now questioning the basic premise of US research policy is unsettling.

Brown would like his report to prompt a fierce debate of the issues. Running for reelection in the autumn in what is expected to be a close race, Brown hopes his strong words will at least inspire his successors if he cannot attempt to legislate such radical changes himself. The test of that strategy will begin on 24 September at the first of a series of hearings before the committee.

Christopher Anderson

* Report of the Task Force on the Health of Research, Committee on Science, Space and Technology, US Congress, 1992.

NEJM restricts use of SI units

Washington. Responding to the wishes of its readers, the *New England Journal of Medicine* (NEJM) has dropped its preference for Système Internationale (SI) units in scientific measurements. Beginning in July, units more common in the United States became the standard rather than appearing only upon the request of the author. Meanwhile, SI units have been relegated to parentheses — and have almost disappeared from tables and charts. One major shift was to express cholesterol levels in milligrams per decilitre instead of millimoles per litre.

The move reverses a decision made several years ago by NEJM and many other US medical journals to emphasize SI units because they are more informative and also more commonly used. But US physicians, commercial laboratories and hospitals have resisted the switch, and the NEJM concluded that it was "unwise to publish in one

system while most of our readers use the other system in their practices."

So far, no other journal has followed suit. For instance, the American Medical Association says it is sticking with SI units because the readership of its journals is evenly divided between the United States and the rest of the world.

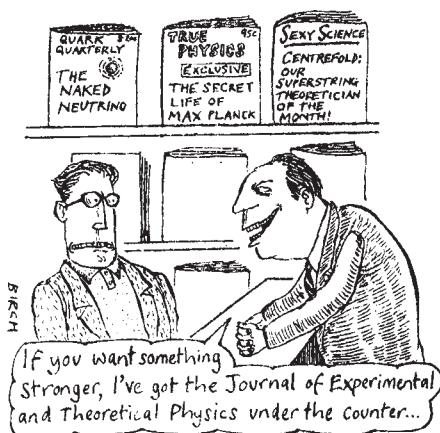
NEJM's retreat has outraged some, especially in Europe. An editorial last month in the *British Medical Journal* compared the NEJM's decision to "the World Health Organization deciding that trying to eradicate polio is just too tiring".

Edward Campion, an NEJM deputy editor and author of the editorial announcing the decision, called the controversy "a tempest in a teapot". He said the change affects only "which [unit] comes first in the parentheses."

Traci Watson

Russian academy makes deal with US company on journals

Moscow. At a time when courage is needed to invest in the Russian economy, two US organizations are actively competing for the rights to publish hundreds of journals and magazines from the Russian Academy of Sciences. Pleiades Publishing, an offshoot of General Media Inc. (GMI), which publishes *Penthouse* and *Omni* magazines, appears to hold the upper hand in the bidding war, but the American Institute of Physics (AIP) has retained its rights to several well-known journals and hopes to add to its stable



in forthcoming negotiations. Both organizations say that they hope to profit commercially as well as providing a service to international science.

The Russian academy needs outside investors to continue publishing more than 200 journals and specialty publications. Operating costs were once met by its predecessor, the former Soviet Academy of Sciences, but the Russian academy has been unable to pay salaries, much less to purchase needed equipment, since assuming control last year.

That is why academy officials reached an agreement this month with Pleiades to create a successor to Nauka, which now publishes and distributes its journals. General Media, the minority owner but chief source of capital in the venture, has promised to spend several million dollars to upgrade the academy's publishing facilities as well as to publish editions in both Russian and English of an unspecified number of journals and magazines. In return, the company receives 49 per cent of the revenues.

"We've been doing business in Russia for three years", says John Evans, president of GMI International. "We've invested in the market and it offers potential profits, not just from this venture but from other forms of translation that may evolve."

The agreement has made AIP work harder for its slice of the Russian publishing pie. In

1988, the institute signed a five-year contract with the Soviet copyright agency overseeing all translation agreements to publish English-language versions of 19 scientific journals, including the *Journal of Experimental and Theoretical Physics* (JETP), *JETP Letters* and *Uspekhi Fiz.* It required AIP to pay royalties, amounting to a little more than \$1 million annually, to be passed to authors.

However, that agency no longer exists and no successor has been formed. As a result, AIP is withholding its 1991 royalties cheque until "we can pay it in a way that will benefit the journals", says AIP's executive director, Kenneth Ford.

In the meantime, AIP has signed up six academy-owned journals and two published by independent institutes in Dubna and Kharkov, Ukraine, and hopes to reach agreements with six others that have expressed an interest in remaining with AIP. The contracts increase the size of its annual payments and include signing bonuses and advances against future royalties to help the Russian journals survive their present financial crisis. Although the academy intends its agreement with Pleiades to cover all its publications, there is a clause that permits journals "with a long history of working with foreign, nonprofit scientific publishers" to make their own agreements.

Evans estimates that "perhaps 100" of the more than 200 academy journals are commercially viable but that it is impossible to know how well any journal will do in Russia because a free market has never existed. "We can turn a profit with 60–70 subscribers", he says about journals with annual subscription prices ranging from \$600 to \$1,600. AIP says that its goal is to provide the English-speaking world with journals of high scientific quality as well as to be commercially successful. Both Evans and Ford say that international readers have gained from having Russian journals published by their organizations. Each cites faster publication times — in some cases, almost simultaneously with the Russian version — better translations and a more professional appearance.

The academy has been criticized for doing business with a company that also publishes an adult entertainment magazine. Evans says that there is no connection between Pleiades and *Penthouse* and that GMI's other businesses are irrelevant to the Russian venture. But Ford says that the Russian academy should not be doing business with GMI, which he says "is the publisher of trash", adding that he is "saddened that the academy found it necessary to enter into this relationship".

Vladimir Pokrowsky

ESA drops Hermes and turns towards links with Russia

Munich. Europe's beleaguered attempts to maintain an independent space programme in the face of recession moved forward last week when the European Space Agency (ESA) agreed to forge closer links with Russia while cutting back on its own projects. The agreement, expected to be approved at the ESA ministerial meeting in November, ends a period of doubt about Europe's ability to reach a consensus on space.

ESA's troubles started last year when Germany found itself unable to meet its financial contribution and other member states were unwilling to make up the shortfall. Since then, ESA has reconsidered its ambitious and increasingly expensive long-term plans, including participation in the launch of a manned space plane, Hermes, by the end of the century. The fate of the US\$7.5 billion Hermes programme hung in the balance until June, when its abandonment was recommended as part of a cost-saving package linked to plans to seek other partners (see *Nature* 357, 351; 1992).

Last week, ESA announced that negotiations between its member states had resulted in reductions of an additional ECU1 billion (US\$1.4 billion) from the 1993–2000 budget proposal, which now stands at ECU22 billion. Half of this saving will occur between 1993 and 1995, including ECU300 million saved by abandoning an unmanned demonstrator, known as X-2000, that was expected to attract international partners.

But ESA now believes that its future lies with the Russians. Jean-Jacques Dordain, counsellor to ESA's director general, says that negotiations with the cash-strapped Russians are no longer confined to the purchase of Russian hardware and the use of Russian facilities; both sides are seriously considering genuine joint ventures for developing space vehicles. But he says ESA will not put all its eggs in one basket. "We will still continue our own key technology work so that we can either share project development on equal terms with the Russians or — if the Russian deals fall through — we will still have the knowhow to develop an autonomous space vehicle."

The Germans, who have led the way in strengthening links with Russian space science, are convinced that cooperation is both achievable and productive. Last month they started a programme to save Russian space science during the next few uncertain years. Research minister Heinz Riesenhuber has pledged DM20 million (US\$15 million) a year, until 1996, to keep Russian space science and technology afloat while officials explore what Russia can offer.

Alison Abbott

Japanese scientists begin closed debate on size of contribution to SSC

Tokyo. The obscure process by which Japan will decide whether to invest as much as \$2 billion in the US Superconducting Super Collider (SSC) lurched forward last week as a working group of the Council of Science and Technology, Japan's highest science policy-making body that is chaired by the prime minister, began closed-door hearings on the issue. But it seems that those in charge have no better idea about the likely outcome than the rest of the world.

There are several government-related bodies involved in the process, which is expected to be completed by the end of the year. The working group of the Council of Science and Technology headed by Wataru Mori, a former president of Tokyo University, heard testimony from a handful of scientists representing all sides of the issue. At the same time, a group of US and Japanese bureaucrats meeting as part of an agreement reached early this year by the Japanese prime minister, Kiichi Miyazawa, and the US president, George Bush, is looking into the technical aspects of collaboration.

Mori and Jiro Kondo, president of the Science Council of Japan and a member of the working group, would not speculate on the outcome of their deliberations. Mori

does not know even if the recommendations of the scientists will be combined with those of the international working group before they are submitted to the prime minister.

The majority of Japanese scientists are opposed to joining the SSC under the terms put forward by the United States, namely, that Japan must contribute a substantial amount to its construction. But the bureaucrats are more likely to be influenced by the fact that Bush has endorsed the project and has personally urged Miyazawa to do likewise.

Akira Itoh, director of superconductivity in the Agency of Industrial Science and Technology of the Ministry of International Trade and Industry, who serves on the US-Japan working group, says that the majority of Japanese high-energy physicists want to participate in the SSC project. But in fact that support comes from only a minority of the country's high-energy physicists, mostly senior scientists who will retire before the proton-proton collider is built (see *Nature* 358, 266; 1992).

The science council, some 210 academics elected by scientific societies to advise the government, debated the SSC at its general assembly last year but reached no

conclusion. Kondo, wishing to act before Bush visited the country in January, wrote his own unfavourable recommendation to the prime minister after clearing it with one of the council's steering committees.

Kondo says that his recommendations, presented in person to Miyazawa, say that the SSC is an important tool for basic science but that Japan must first improve its own domestic research. Kondo says Miyazawa thanked him for his recommendations and said that Kondo's comments are the first advice on the SSC that he has received from Japanese scientists.

The recent visits to Japan by officials of the European Laboratory of Particle Physics (CERN) seeking Japanese support for the planned European Large Hadron Collider (LHC) (see *Nature* 359, 5; 1992) has prompted other senior Japanese university professors to speak their minds on the SSC. One adviser to the Science and Technology Agency and the Ministry of Education, Science and Culture, says he worries that the SSC issue may tear apart the Japanese scientific community. He says that if Japan contributes to the SSC for political reasons, it should do so in a way unrelated to science.

David Swinbanks

China juggles economic growth and basic research

Beijing. A new five-year programme to strengthen China's economy is the latest example of a struggle by scientists around the world to preserve adequate funding for basic research.

The initiative, called the Climbing Programme, was conceived by the State Science and Technology Commission with support from the country's State Education Commission, the National Natural Science Foundation and the Chinese Academy of Sciences. Each of 30 projects will receive 1 million yuan (US\$200,000) a year for five years, and the overall programme will involve some 6,000 scientists and engineers.

On the surface, the programme represents a major commitment to basic research. Each project satisfies at least one of four criteria: it represents frontier research; it contributes to China's economic development; it draws on the country's unique natural resources, geography or cultural heritage; and it advances a field in which China is already pre-eminent. The projects were chosen by a 45-member committee of scientists and science administrators led by Zhang Chun-Hao, director of the science foundation.

Ten of the projects cover work in such rapidly expanding fields as high-tempera-

ture superconductivity and molecular biology; eight projects, including those involving nanomanufacturing and high-yield, highly resistant crops, are expected to have important commercial applications. Another eight examine the ecosystem of the Qinghai-Tibetan plateau, the world's youngest mountain range, and Jingluo, a traditional approach to medical science, for example, while the remainder cover areas, including mathematical proofs by computer, in which Chinese scientists are recognized as world leaders.

Yet basic research remains on a precarious footing in China, which invests only 0.7 per cent of its gross national product on research, far below the average for the industrial world. The president of the Chinese Academy of Sciences, Zhou Guang-Zhao, believes that one-third of the academy staff should leave their laboratories and take up applied work in industry. Just last month, he told a national symposium on production and research that "as long as a high-quality research community is maintained, China's basic research will not be affected by encouraging researchers to enter the main battlefield of economic development".

There is also considerable unhappiness

over how the final round of proposals was selected for the Climbing Programme. With 172 projects competing for 18 remaining slots, the selection committee gave outstanding scores to 25. However, government officials could not be persuaded to expand its quota of winners, and seven worthy projects were arbitrarily dropped. In addition, the National Natural Science Foundation approves only one in four applications, despite an estimate by officials that almost half of them are excellent proposals that deserve funding.

This unhappiness surfaced unexpectedly during a planning meeting for the Climbing Programme. In a breach of Chinese etiquette, Wang Gan-Chang, a prominent member of the Chinese academy and a co-founder of a high-technology project launched in 1986, confronted Song Jian, director of the State Science and Technology Commission, with the accusation, "Where were you earlier?". Song replied that what matters most is "the present".

Despite these concerns, the government has repeatedly said that it intends to reward innovative and first-class science. The Climbing Programme is seen as one step on that road to excellence.

You Qin Li

Flood of applicants delays EC training grants

Munich. A new European Communities (EC) programme intended to promote mobility for research workers within Europe is again being criticized. Overwhelmed by the number of applications, EC officials decided last week to defer processing them, annoying researchers who say they were given an unreasonably short deadline to submit very complicated proposals.

The Human Capital and Mobility Programme has confused and frustrated scientists since its launch earlier this summer (see *Nature* 358, 98; 1992). The ECU488 million (US\$710 million) programme is an ambitious attempt to make European science as cohesive and mobile as US science. Most of the money goes to transnational postdoctoral fellowships; 45 per cent of it to individual host institutions offering postdoctoral projects to foreign EC

scientists and 40 per cent to networks.

The networks programme has attracted particular criticism. To qualify, researchers must establish links with at least five research laboratories in at least three EC countries. When the early August submission date for applications was finally announced in July, applicants complained that they needed more time. The EC justified its haste by saying that the money would be lost if not allocated this year and that several months would be required for processing.

Despite the complexities, more than 3,000 applications were received and queues ringed the Brussels-based offices on the closing date. The strictness of the deadline was relaxed only for applications from Portugal; they were flown to Brussels by courier on deadline but were lost by the airline and arrived three days late.

The EC, whose support system was capable of handling as many as 500 applications, realized it would not be able to cope with the unexpectedly heavy load. It decided to defer processing of network research project applications and to concentrate instead on the identification of host institutes. Most of these decisions have now been made.

Although the programme's management committees do not meet until November, a spokesman for the EC says that it may still have time to allocate all the money. In any case, the August deadline is not the only opportunity for scientists to receive Human Mobility funds. Unlike most other EC programmes, there is a rolling deadline for applications until 1994. Next year the EC will award grants three times, in March, June and November, from a budget of ECU260 million.

Alison Abbott

Newton statue under fire

London. A 12-foot-high bronze statue of Sir Isaac Newton commissioned for the forecourt of the British Library's new headquarters under construction in London has caused a row between the library and scientists. The statue, seen as a maquette below, is an interpretation by Sir Eduardo Paolozzi of a woodcut by the eighteenth-century mystic William Blake, right.

As chronicled in *The Times*, the first salvo was a complaint by Blake enthusiasts that the artist's vision had been misrepresented. Their letters drew a response from the building's architect, Colin St John Wilson, who said that Blake's view of Newton and his disdain for scientific obsession with the measurable was retained in the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

sculpture. "This equivocal attitude to the values of science is shared by many eminent scientists as well as laymen", Wilson wrote. A few days later, Paolozzi defended his interpretation.

Wilson's comment has infuriated scientists such as Mansel Davies, emeritus professor of chemistry at the University of Wales, Aberystwyth, who feel that such a strong anti-scientific motif is inappropriate for a building that unites for the first time the library's science and humanities collections. "It represents an antagonism to the mathematical basis of science that the Council of the British Library has accepted without question," he says.

The statue, expected to cost £175,000, is not yet cast. The debate continues.

Ian Mundell

New life planned for British plasma science

London. The Science and Engineering Research Council (SERC) and industry are considering initiatives to revitalize British plasma research. The intention is to bridge the gap between fusion research, which has seen a dip in funding in recent years, and other areas of plasma science, particularly those with industrial applications. Although glad of the attention that the wider field of plasma science is now getting, some academic researchers are concerned about placing their future entirely in the hands of industry.

Fusion energy programmes have long dominated plasma research in Britain, with the largest groups at Atomic Energy Authority's (AEA) Culham laboratory, home of the Joint European Torus (JET), and at Imperial College London. AEA now spends £10.8 million (US\$21 million) on fusion research, of which £400,000 goes to universities. SERC spends about £250,000 annually on plasma research related to fusion.

However, organizational changes at the AEA have left some groups stranded. The mechanisms to take up the slack in supporting work peripheral to fusion, but considered worthy of study in its own right, have never really been established. Similarly, there has been no increase in support for the study

of non-equilibrium and impure plasmas, fields that have grown in recent years.

"[The level of] support for plasma physics in Britain is sub-critical", says David Burgess of Imperial College. "There are good, strong groups that cannot move from one area to another."

The SERC is considering a plasma science initiative, proposed by Burgess, to cover plasmas with interesting technological features as well as areas spun off from the fusion studies. At the same time, researchers emphasize that such applications are often derived from fundamental work, which cannot be allowed to wither away.

The technologies in question, identified in a report from the independent Centre for the Exploitation of Science and Technology (CEST)*, include using plasmas in manufacturing, particularly in modifying surfaces, and in environmental clean-up, where they can be used to destroy low levels of toxic waste. But industry needs better methods for modelling plasmas before it can scale up experimental procedures.

The report says that more needs to be done to link industry and university scientists, to increase awareness of what plasmas can do and to transfer skills. Although researchers are keen for this sort

of exchange, they do not want to sacrifice their independence.

"I can understand why CEST regards these areas as so important, but you have to distinguish those that industry should pay for and those that, as part of the scientific future of this country, are down to the SERC", says Burgess. He said that industry must pay the going rate for whatever it wants that is now done at universities. **Ian Mundell**

*Industrial plasmas: focusing UK skills on global opportunities (£30), CEST, 5 Berners Road, London N1 0PW (tel. 071-354-9942).

Big projects eat up NASA's future budgets

Washington. Despite the recent shift of the National Aeronautics and Space Administration (NASA) towards smaller and cheaper missions, nearly all the agency's budget is already committed to large projects that would be politically difficult to kill, accord-

University wants to buy journals directly

London. In an attempt to stem the rising cost of scientific journals, the University of Kent at Canterbury is considering creating its own subscription agency as a means of getting more for its money. By buying journals direct from publishers, the library and university departments can take advantage of the agency discount offered by some publishers and avoid the mark-up, or service charge, usually added to subscriptions by commercial agencies. University administrators are discussing with publishers the viability of such a scheme.

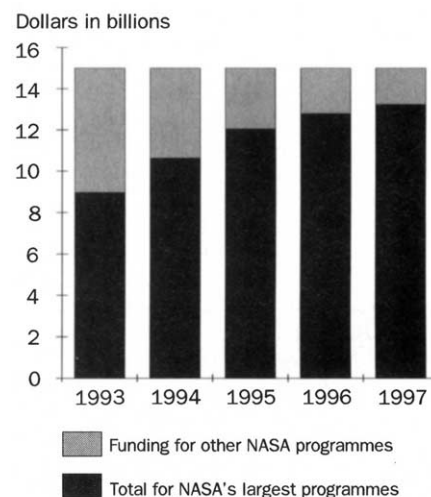
Cyril Isenberg, a Kent physicist, says that it should be possible to save tens of thousands of pounds despite the costs involved in running the office. The university now spends £300,000 a year on journals.

However, it may be that Kent has underestimated the costs and manpower involved. The agencies operate on a narrow margin, earning a discount from publishers for taking on the burden of dealing with a multitude of customers. Although the agencies expect to earn a profit by practising economies of scale, some of the smaller operations are said to be struggling. The universities are charged for the service, which gives them the advantage of buying from one source and a certain amount of credit.

Journal publishers appreciate the problems that inflation and rising subscription prices have created for university libraries and encourage any efforts by libraries to improve the situation. However, they are unwilling to anger the agencies, which hold a very powerful place in the publishing chain, and privately they doubt whether university agencies would be offered the same discount.

Meanwhile, other British universities have attacked the problem by working within the existing system. Both the colleges of the University of London and a group of universities in the south of England have bulk-order arrangements with the agencies that offer them a greater discount. When cancellations have been unavoidable, obtaining individual papers through special databases has helped them to fill the gaps.

Although the British Library's document supply centre in Boston Spa dominates the British market, such services are of increasing importance to libraries. B.H. Blackwell, a journal supply agency based in Oxford, has just announced an agreement with a US company, CARL Systems, that will make available a commercial document supply service derived from a base of 15,000 to 20,000 titles. **Ian Mundell**



ing to a congressional analysis released last week. At current levels, projects such as the space shuttle, the space station and the Earth observing system will consume nearly 90 per cent of NASA's funding by 1997, leaving little money for a new crop of smaller missions such as a proposed Pluto probe (see *Nature* 358, 701; 1992).

Auditors from the congressional General Accounting Office (GAO) assume that NASA's budget will remain near its current level of \$15 billion for the next five years, a projection that seems increasingly realistic, and they criticize NASA for ignoring this fact in its five-year plan. Agency officials say that they are working on a new plan that would lower the cost of the large programmes.

Christopher Anderson

Hungarian law will protect academics strapped for cash

Budapest. The rights of researchers, students and teachers in universities and research institutes to work with total political and ideological freedom will be written into Hungary's new constitution this autumn.

The new laws will formalize the quiet research revolution that has been going on for the past few years (see *Nature* 352, 745; 1991). They provide a silver lining to continuing clouds of economic misfortune and follow several years of steady progress towards westernization of research organization after the fall of Hungary's 40-year-old communist regime.

The reforms, outlined in the 'Act of the Academy', restore the autonomy of the Hungarian Academy of Sciences, as well as restructuring its 39 institutes. They are based on the efforts of Domokos Kosáry, who spent four years in jail as a communist dissident and who was elected president of the academy in 1990. The academy, founded in 1825, runs most of Hungary's independent research institutes as well as looking after the interests of national science.

Many of the reforms are already under way. More than 200 scientists, half from universities, are analysing the recent performance of the academy's institutes as members of 40 evaluation committees. This month, teams from the International Council of Scientific Unions (ICSU) will assess three of the larger institutes. Their findings will be used to determine how next year's budget is allocated.

Resources are now allocated by the Hungarian Research Fund, known as OTKA, under a peer-review system established in 1986. OTKA is rethinking its original deci-

sion to distribute its relatively small fund fairly evenly between successful applicants and is preparing to give proportionately higher rewards for excellence.

The Higher Education Act — the first comprehensive legislation in Hungary's 600-year history of higher education — is also expected to be passed this autumn. It provides autonomy for universities and reestablishes their rights, removed during the communist years, to conduct postgraduate research, award PhDs and carry out the process of 'habilitation' (tenure). Until now, two parallel systems of research-based doctorates have existed: the two-year 'university doctor', awarded by the universities, and the 'candidate of science', awarded by a now-defunct government committee often accused of abusing its political power. In their place will be a single PhD programme similar to those in the West.

These new postgraduate activities will be monitored by an accreditation committee made up of Hungarian and foreign academic researchers and experts from outside higher education. One of the committee's first tasks will be to assess the country's 30 universities in the same way that the research institutes were judged.

But rising political fortunes must contend with a failing economy. This year's research budget, frozen at HUF4.7 billion (US\$60 million), is being devoured by an inflation rate of 30 per cent. According to István Láng, general secretary of the Hungarian academy, research has done better than other government programmes but inadequate funding makes it essential to pick the best research.

Alison Abbott

Hard times for high- T_c press

Nikkei Superconductors, a twice-monthly Japanese newsletter created during the 1987 boom in high-temperature superconductors, will cease publication later this month. It joins at least four other superconductivity publications that have folded or cut back recently, victims of a slowing of the rate of scientific progress in the field.

In the United States, *Supercurrents* has closed and *Superconductivity News*, after a string of financial problems, has missed several issues. Its editors are considering reconfiguring the newsletter as a one- or two-page weekly. For the moment, that leaves just *High T_c Update*, a government-subsidized newsletter, and the commercial publications *Superconductor Week* and the quarterly *Superconductor Industry*. Two international superconductivity newsletters

published in Britain by Elsevier are still in operation, although the company says that circulation is down by about 20 per cent. And two journals funded by the British government — *Superconductivity Titles* and *Superconductor Focus* — were terminated last December as part of an overall belt-tightening.

Kevin Ott, director of the Washington-based Council on Superconductivity for American Competitiveness, attributes the decline in publications to an evolution in the field towards commercialization. Although progress in basic research has slowed, interest in the application of superconductors is increasing. Attendance at last month's biennial Applied Superconductivity Conference in Chicago was 25 per cent higher than in 1990, at a record 1,500. **From our bureaus**

NEWS IN BRIEF

Munich. The heaviest known chemical elements — with half-lives of around 5 ms. — were finally named last week by scientists at the Institute of Heavy Ion Research in Darmstadt, Germany, where they were created by fusion reactions in the early 1980s. Element 107 is to be called nielsbohrium, after the atomic physicist Niels Bohr. Element 108, a



Niels Bohr

fusion product of Pb208 and Fe58, is now hassium, named for the German state of Hessen, which supports the institute financially. Element 109 — meitnerium — is a tribute to Lise Meitner, whose theoretical

interpretation of Otto Hahn's experiments was fundamental to the discovery of nuclear fission.

Unlike some previous attempts to name transuranic elements, this terminology is expected to be accepted by the international scientific community. During the cold war, element 104 was discovered concurrently in the Soviet Union and the United States, but debate continues on whether its name should be kurchatovium or rutherfordium, after the leading English and Soviet nuclear physicists of their time. **A.A.**

Washington. Seventy-one days after he received the first baboon-to-human liver transplant, a 35-year-old man died last week at the University of Pittsburgh. After his death of a haemorrhage in the brain, the university disclosed that the patient had been HIV-positive, although he had not yet begun to exhibit AIDS-defining symptoms. University officials said that the condition did not appear to have contributed to his death.

The patient, whose identity has been withheld, was considered to be a poor candidate for a human liver because his own liver had been destroyed by hepatitis B, which would have presumably attacked any human transplant. Rather than depleting a supply of human organs that is already too small to meet demand, the surgeons decided to use the liver of a baboon, which is thought to be immune to hepatitis B. The Pittsburgh surgeons said that an autopsy revealed only mild evidence of rejection, adding that they will continue with plans to transplant three more baboon livers into humans. **C.A.**

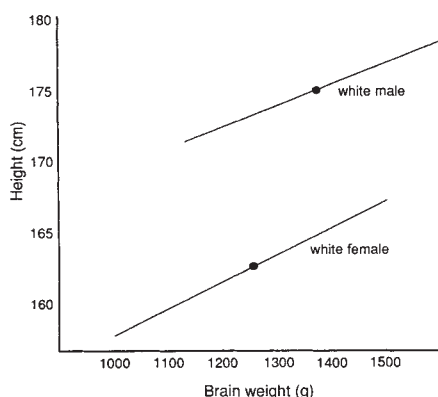
Beijing. The National Natural Science Foundation of China, created in 1986, has received an increase of 28 per cent in its budget for 1992. The foundation has yet to decide whether to use the extra money to increase the size of the average grant or to make additional awards. **Y.Q.L.**

Brain size differences

SIR — Recent claims by Rushton and Ankney of racial and sex differences in brain size¹⁻³ are accounted for by a simple artefact of the statistical method employed. I illustrate this effect below.

I focus on the claim that men have larger brains than women, because this difference was the most dramatic of all comparisons made¹. I use the autopsy measurement of brain weight and body dimensions of Cleveland adults presented by Ho *et al.*^{4,5}. These same data were recently used by Ankney⁶ to bolster the conclusion that men have larger brains than women, once body size differences between men and women are accounted for. This conclusion is currently receiving a great deal of attention in the Canadian press⁷. Ho *et al.*^{4,5} did not present the raw data, but all necessary quantities can be computed from their tabular summaries. Using linear regression⁸ of brain weight on body size, men and women of the same body size are compared. The men have larger brains on average. For example, mean brain weight of white men 170 cm tall is about 100 g greater than of white women of the same height^{1,6}. A similar result holds when body weight is used instead⁶.

Do men have larger brains? The flaw in such a claim is forcefully illustrated by a further analysis of these same data. Rather than compare brain sizes between men and women of similar height, I compared mean heights of men and women having the same brain size. If men truly have larger brains for their body size than women, then men should be shorter than women of equal brain weight. Yet the opposite is true: white men are more than 10 cm taller on average than white women with the same brain weight (see figure). Indeed,



Linear regressions of body height on brain weight in white men and women. Regression statistics were computed from data in Ho *et al.*^{4,5} For men, $y = 0.015Y + 154.4$ ($n = 414$). for women, $y = 0.19Y + 138.7$ ($n = 388$). Lines extend two standard deviations to either side of mean brain weight, indicated by the large points.

to find a male brain-size category in which mean height is equal to that of women with 1,500 g brains, it is necessary to extrapolate downward to a male brain of about 850 g. Clearly, by this criterion women have much larger brains than men. The same is true when body weight rather than height is used. Furthermore, racial differences disappear when the data are analysed in the same way.

The source of the conflicting conclusions is the well-known 'regression effect'⁸. This paradox occurs because brain size of individuals varies partly independently of body size, and vice versa. The effect is enhanced in the present case because the causes of size variation between individuals of the same sex (which may be chiefly environmental) are not the same as the causes of differences between the sexes (which may be largely genetic). Hence it makes little sense to use natural variation in brain and body size within sexes to correct for differences between men and women.

Dolph Schluter

Department of Zoology,
University of British Columbia,
Vancouver, BC, V6T 1Z4 Canada

SIR — In your comment on Rushton's recent data showing different mean cranial capacities among Asian, Caucasian and Afro-American US Army personnel¹, you suggest that Caucasian military personnel may be of relatively higher quality, compared to the Caucasian population as a whole, than Afro-American, thus producing a biased estimate of the difference in brain size.

This is surely improbable. The US Army administers aptitude tests to potential recruits and accepts only those who score above a minimum threshold. Afro-Americans score lower on aptitude tests than Caucasians, so a greater proportion of Afro-Americans in the lower ability range are screened out by the tests. The admission tests appear to screen out 3.4 per cent of Caucasians and 30 per cent of Afro-Americans⁹. This procedure raises the mean ability levels of both Caucasians and Afro-Americans in the US Army, as compared with the general population, but it raises the mean ability of Afro-American military personnel considerably more than that of Caucasians. The effect is to raise the mean IQ of Caucasians in the US Army by 1 IQ point (100 to 101) and of Afro-Americans by 6 IQ points (85 to 91).

The effect of military selection tests in reducing the black-white difference in the general population goes some way

towards explaining why the black-white differences for cranial capacity reported by Rushton are much lower than those obtained by Beals, Smith and Dodd¹⁰. Their study of worldwide data on approximately 20,000 crania found that Caucasians have a mean cranial capacity of 1,362 cm³, indistinguishable from Rushton's figure of 1,361 cm³ for American Caucasian military personnel. But the Beals *et al.* figure for the cranial capacity of Africans is 1,276 cm³, considerably lower than Rushton's figure of 1,346 cm³ for Afro-Americans in the US Army. The most probable explanation of these differences is that the selection procedures for the US Army exert only minimum bias on the intake of whites but screen out large numbers of low-ability blacks. Thus you are correct in suggesting that there is a bias in Rushton's military sample, but in the opposite direction from that suggested in your comment.

Richard Lynn

Psychology Department,
University of Ulster,
Coleraine,
Northern Ireland BT52 1SA, UK

SIR — Science does not exist in a vacuum and is therefore subject to all the political, social or economic influences that exist in this world. Political correctness has its place, whether scientists like it or not. Everything relating to science, that is, the method, the data, the interpretations and results, are all as subjective as any other human endeavour (simply because it is a human endeavour). Thus, science and researchers are not immune to attacks from a nonscientific viewpoint such as the politically correct.

The conclusions drawn by Rushton² and Ankney³ in the brain size/IQ debate are dubious at best. Mary Warnock, commenting on John Stuart Mill's essay 'The Subjection of Women', writes of Mills: "He was not prepared to accept any argument purporting to show that women were *naturally* inferior in intellect or originality to men. For, he said, there had never been a chance to test such a hypothesis. Only after generations of equal education could any proposition about the powers of women compared with those of men be considered."¹¹ In other words, neither Rushton nor Ankney considered the political, social and economic differences between the sexes. If men and women were educated in an environment where both sexes were equally encouraged to achieve, given equal opportunity, given equal rights and there existed no discrimination of any kind, both sexes would no doubt excel equally in all attributes to mental abilities. If the brain size data are true, this means that women's brains

work more efficiently than men's. Also, isn't it possible that the difference in brain size (if any) is the result of generations and generations of oppression of women by men? Has either Rushton or Ankney considered other reasons (or *all* the other possible reasons for that matter) for brain size difference that may be just as statistically significant as that from gender alone? What every experiment needs is a control. For the brain size/IQ debate, this appears impossible and thus Rushton's and Ankney's conclusions are without merit.

F. C. S. Tsai

Department of Chemistry,
University of British Columbia,
Vancouver, BC V6T 1Z1,
Canada

SIR — Professor J. Phillippe Rushton offers an analysis of the cranial capacities of the several ethnic groups recruited by the US Army, with an additional comparison between those of men and women. John Maddox¹ has been rightly sarcastic about the biases implicit in Rushton's study: if ethnic differences exist, are they based upon neurological factors that have never been examined? How tightly are neurons packed within brains? Are differences of myelination involved? In the context of intelligence (whatever that may mean), how do the relevant neural networks function, anyway? For serious neuroscientists, much is uncertain. On a sociological front, how are men of diverse ethnic origin recruited by armies in the first place? What personal needs drive them?

Rushton's enormous study seems to have been based on anthropometric data gathered to help contractors to the US Army to provide a range of helmets. Should we take it seriously? Surely not.

Were it possible to accept Rushton's efforts as methodologically sound, we could leave him alone to get on with it. Alas, his record is not impeccable.

In his paper¹² "Genetic similarity in male friendships", he used data derived from application of a test of 'conservatism' invented in England in the late 1960s by G. D. Wilson and J. R. Patterson¹³ This farrago devised for the

British *milieu* of the period was transferred to Canada. To what were Rushton's harmless subjects asked to respond? A small sample will suffice: Self-denial, evolution theory, school uniform, hippies, sabbath observance, patriotism, modern art, colonial immigration, Bible truth, pajama parties, inborn conscience. Surely, enough is enough.

A. David Blest

Developmental Neurobiology Group,
Research School of Biological Sciences,
Australian National University,
Canberra, Australia

Anal sex

SIR — John Maddox (*Nature* 358, 13; 1992) criticized *The Sunday Times* for reporting that the Birmingham haemophilic who infected four women with human immunodeficiency virus (HIV) had had anal intercourse with at least three of them. Neville Hodgkinson in his reply (358, 447; 1992) asked "Why is a journal dedicated to science so afraid of facts?"

I would like to put the same question not only to *Nature*, but also to those involved in AIDS research, and especially to those responsible for the information to the public about risk factors in heterosexual transmission of HIV. The reason is as follows.

Maddox pointed out that the extra hazard of anal intercourse is not novel. That is true. It has been known for at least seven years within the AIDS scientific community. But is it also well-known outside this community? I doubt it. At least in Sweden I have never seen government sponsored campaign indicating that anal sex is risky sexual behaviour.

Maddox refers to a report published by the European Study Group on Heterosexual Transmission of HIV (*British Medical Journal* 304, 811; 1992). In that report, the relative risk of HIV transmission from men to women is increased by a factor of 5.1 in anal intercourse, compared to vaginal intercourse.

What does this relative risk factor mean in practice? Raw data show that 46 per cent of women who had had anal intercourse with HIV-infected men became HIV-positive. This is most alarming to me (and probably to most other people) but not for the European Study Group. They conclude that this high-risk sexual practice (anal sex) was not essential for transmission as 40 out of the 82 infected women never practised anal sex (the other 42 did). Evidently they prefer to talk about relative risks obtained by complex statistical methods and miss or hide the most important message from facts.

A major conclusion from the report in

question should be that a dominant risk for transmitting HIV from an infected man to a woman is to practise anal sex. This is also supported by common sense (which sometimes is of value also in science). Male homosexuals frequently practise anal sex. Everybody knows about the high frequency of HIV transmission within that group.

That anal sex is practised by the heterosexual and bisexual population has to be recognized by society. In the two study groups in the report, the frequency was 23.7–31.8 per cent which is in accordance with results from the Stockholm area. Thus scientists dealing with AIDS research and people, organizations and media responsible for giving information and advice on AIDS protection have a duty to ensure that the following facts are clearly stated to the public: (1) anal sex is practised by many heterosexuals and bisexuals; and (2) anal sex is relatively more dangerous.

This information is especially important for women, because they are the main victims.

Per-Erik Åsard

Department of Hospital Physics,
Danderyd Hospital,
S-182 38 Danderyd, Sweden

Sex and gender

SIR — According to Fowler's *Modern English Usage* (Oxford University Press, 2nd edn 1965, revised 1983), "GENDER n. is a grammatical term only. To talk of persons or creatures of the masculine or feminine g., meaning of the male of female sex, is either a jocularity (permissible or not according to context) or a blunder."

For reasons of political correctness rather than biological delicacy, however, the nonjocular use of gender as a euphemism for sex seems now to be getting established. Whether this really serves any useful purpose for human sex is perhaps arguable, but its application to nonhuman animals should surely be resisted.

An example appears in *Nature* 358, 704 (1992), on the recovery of an enormous fossil of the ornithischian dinosaur *Stegosaurus*, found recently in Colorado. This has much larger back plates than in another specimen near by, "which may suggest that plate size is an indication of gender, and that the specimen is male".

But according to the rules of zoological nomenclature, and of Latin grammar, the *gender* of the name *Stegosaurus* must be masculine. It is the sex of this particular specimen, whether male or female, which is in question.

C. B. Goodhart

Gonville & Caius College,
Cambridge CB2 1TA, UK

- Maddox, J. *Nature* 358, 187 (1992).
- Rushton, J. P. *Nature* 358, 532 (1992).
- Ankney, C. D. *Nature* 358, 532 (1992).
- Ho, K., Roessmann, U., Straumfjord, J. V. & Monroe, G. *Arch. Pathol. Lab. Med.* 104, 635 (1980).
- Ho, K., Roessmann, U., Straumfjord, J. V. & Monroe, G. *Arch. Pathol. Lab. Med.* 104, 640 (1980).
- Ankney, C. D. *Intelligence* 16, 329 (1992).
- Strauss, S. *Globe and Mail*, Toronto (15 August 1992).
- Freedman, D., Pisani, R., & Purves, R. *Statistics* (Norton, New York, 1978).
- Flynn, J. R. *Race, IQ and Jensen* (Routledge and Kegan Paul, London, 1980).
- Beals, K. L., Smith, C. L. & Dodd, S. M. *Curr. Anthropol.* 25, 301–330 (1984).
- Warnock, M. in Wollstonecraft, M. *A Vindication of the Rights of Women* & Mill, J. S. *The Subjection of Women* (Dent, London, 1985).
- Ethol. Sociobiol. 10, 361–374 (1989).
- Br. J. clin. Psych. 7, 264–269 (1968).

NIH labours to maintain excellence despite tight budget and staffing

For years, the US National Institutes of Health (NIH) has been charmingly described in Washington political circles as the 'crown jewel' in the federal Department of Health and Human Services, a treasure deserving of respect and ever increasing amounts of money to support the biomedical research enterprise that works tirelessly to advance the cause of human health. And for years, despite its 'favoured agency' status within the government, NIH officials and the researchers they support have argued that the enterprise is in jeopardy because the research dollars (no matter how plentiful) have not kept pace with the intellectual opportunities in science.

Nevertheless, funding for NIH has always increased in an odd but treasured ritual in which the president, in his budget message, asks for a small increase and Congress, in a game of one-upmanship, adds to it by a considerable amount. Now, for the first time in two decades, it looks as though the ritual will not play itself out. If NIH is lucky, its total funding increase will come to little

more than 3 per cent (see chart), not enough to keep up with inflation. This result, which will affect the number of grants that NIH can support at universities nationwide, will also put a damper on plans to rejuvenate the institutes' intramural research programmes, which employ close to 4,500 professional scientists along with a nearly equal number of technicians and other laboratory personnel.

All of this comes at a moment when Bernadine Healy, director of NIH, is making one of the more determined efforts in the institutes' history to secure financial stability through political cunning. Healy is the power (and inspiration) behind the controversial 'strategic plan' for NIH which, in its third or fourth rendition, has been dramatically called "Advantage America". The idea is that what is good for NIH is good for the United States, economically speaking, thereby linking what Healy calls "corporate NIH" with the once-popular notion from Detroit that "What's good for General Motors is good for America" (see *Nature*

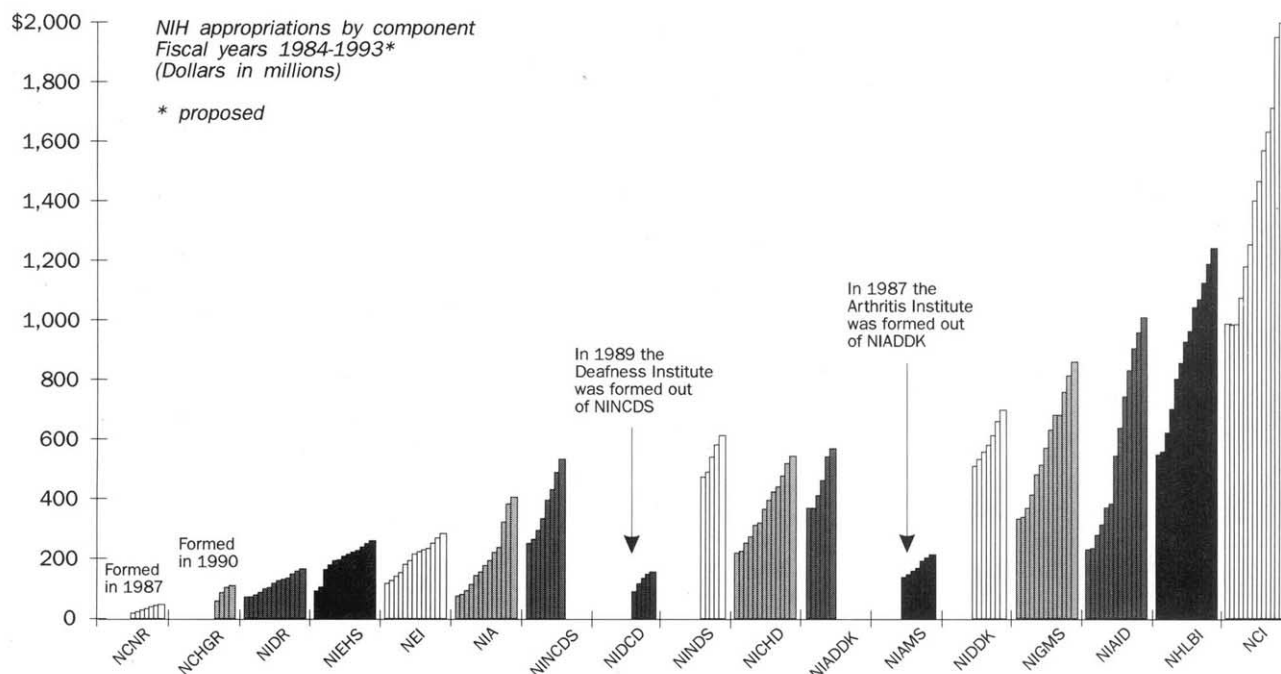
355, 573; 1992; and 358, 355; 1992)

Thus, one of the NIH's newly defined strategic goals (despite protests from many senior scientists) is to "enhance the nation's economic well-being". But for now, at least, the suggestion that the United States' economic picture could be made rosier by spending more money on biomedical research is falling on deaf congressional ears.

Undaunted, NIH continues to think big (and optimistically), one sign being Healy's interest in capitalizing on the institutes' broad intramural strength in basic human molecular genetics and gene therapy (see page 188). To overcome one of NIH's chronic problems in recruiting senior scientific stars (namely salary), she is trying to assure the institutes' stature in this field by offering whoever becomes new head of the National Center for Human Genome Research (replacing James Watson) the two things researchers crave most—space and staff (see page 185).

But for these and other efforts to enhance the intramural research programme, thought

Institutes growth rate varies



NCNR, National Center for Nursing Research; **NCHGR**, National Center for Human Genome Research; **NIDR**, National Institute of Dental Research; **NIEHS**, National Institute of Environmental Health Sciences; **NEI**, National Eye Institute; **NIA**, National Institute on Ageing; **NIDCD**, National Institute on Deafness and Other Communication Disorders; **NINDS**, National Institute of Neurological Disorders and Stroke; **NICHD**, National Institute of Child Health and Human Development; **NIAMS**, National Institute of Arthritis and Musculoskeletal and Skin Diseases; **NIDDK**, National Institute of Diabetes and Digestive and Kidney Diseases; **NIGMS**, National Institute of General Medical Sciences; **NIAID**, National Institute of Allergy and Infectious Diseases; **NHLBI**, National Heart Lung, and Blood Institute; **NCI**, National Cancer Institute.



NIH directors from 1950 through 1989, in order of appearance in office from top left: James Wyngaarden, Donald Fredrickson, Robert Stone. Lower row from left, Robert Marston, James Shannon and William Sebrell, Jr.

in the 1960s and 1970s to be among the very best in the United States, NIH must overcome sizeable problems. One is that its employees cannot escape the hazards of government work. This year, for example, NIH imposed a hiring freeze when it overextended its budget and now must convene a special committee and receive special approval from Healy each time it wishes to hire someone. NIH officials hope the freeze will be lifted by 1 October, the beginning of fiscal year 1993, but for the moment, only postdoctoral researchers and other entry-level positions are being filled.

NIH has also fallen short in its plans to attract more senior researchers. One government-wide initiative, known as the Senior Biomedical Research Service (SBRS), created a category of senior research positions exempt from the usual salary caps. NIH has been allocated 230 SBRS positions and will get even more next month when it absorbs the National Institute of Mental Health, the National Institute on Drug Abuse and the National Institute on Alcohol and Alcohol Abuse. But although the SBRS became law in November 1990, the White House Office of Management and Budget opposes the programme and has delayed its implementation. With the death in February 1991 of its champion in Congress, Silvio Conte of Massachusetts, prospects for the programme are not good. Likewise, a plan for ten privately endowed chairs free of salary caps has stalled. Although the programme was authorized in 1991, no money was ever appropriated for it.

Ironically, even the recent raises given to scientists in the Senior Executive Service (see page 185) might backfire. Pensions are based on the average of a scientist's three highest-paid years, and NIH officials suspect that some of the recent decline in turnover reflects a decision to delay retirement

until employees have recorded three years with their higher salary. "January 1994 is the scary date", says Stephen Benowitz, director of the NIH division of personnel management, referring to the end of the first three-year cycle.

Underlying that concern is a continuing worry about the 'greying of NIH'. The average age of NIH's scientific workforce has increased by 3.3 years in the last decade because of a combination of trends. The happy news is that senior scientists are staying longer, but the bad news is that younger scientists are leaving sooner.

"The perceived negatives [of work at NIH] — that one doesn't come here and stay — are still true", says David Dichek, a senior staff fellow who joined NIH in 1988. The unfortunate result is that, as the current generation of managers and laboratory heads leaves over the next ten or so years, NIH must either promote young scientists into management before

their time or find better ways of recruiting outside talent.

Given these problems, the idea that NIH's intramural programme can sustain in the 1990s and beyond its unparalleled reputation for scientific energy of a quarter-century ago may be wishful thinking. Part of the reason is that NIH, through its extramural grants and contracts, has given rise to dozens of first-class research institutions that



In 1991 Bernadine Healy became the first woman to head the NIH.

now compete for the best people. But Healy points out that "life in the extramural community is no picnic", and the problems facing all of research are certainly no worse than those confronting NIH.

In the end, a healthy NIH and a healthy national research programme depend on each other. The jewel may have lost a bit of its shine, but none brighter exists in the US research firmament.

Barbara J. Culliton & Christopher Anderson

Job prospects are better than ever (but were never as bad as made out)

Forget the 'brain drain'. Fears that the government was losing its best scientists to industry and academia — an apocalyptic picture painted several years ago in part to raise salaries — have now disappeared. In reality, NIH is attracting more young scientists and losing fewer senior scientists than at any time in the last decade.

The truth may be less useful in winning a bigger budget, but it is a sign of health at the largest US science agency and an indication that several new programmes are beginning to make NIH more attractive to scientists.

Applications to NIH's postdoctoral positions are at their highest levels since the Vietnam war, when thousands of researchers joined NIH's intramural programme to avoid being drafted. Turnover among senior scientists is declining, standing at 7.4 per cent for MDs, for example, a figure comparable to turnover at universities and less than in industry. And half-a-dozen new initiatives — from exemptions to federal salary ceilings to more flexible rules on consulting — have raised morale and, more impor-

tantly, brought NIH salaries in line with (and sometimes above) the US average.

The difference between the NIH intramural programme today and the sorry picture officials painted in the 1980s is part perception and part reality. "We may have oversold the whole brain-drain thing," says Lance Liotta, NIH deputy director for intramural research. Although sounding the alarm did gain NIH some relief from government salary ceilings and more money, it may also have become a self-fulfilling prophecy; right or wrong, NIH acquired a reputation as a place to leave.

But out of that perception has come a commitment to change — new programmes and fewer restrictions if possible; simply trying harder if not. "NIH hasn't had a climate where [managers] were making an effort to recruit senior people", says Bernadine Healy, the NIH director. "The fact that they weren't getting them is tied to the fact that they just weren't trying." Since 1979, when the Senior Scientific Service was created for about 100 leading researchers, NIH has recruited only one person to its ranks.

The rest were promoted the easy way — from within.

Healy hopes that the search for a prominent researcher to replace James Watson as the director of the National Center for Human Genome Research will set the tone. For the right candidate, Healy says she is willing to offer 10,000–20,000 square feet of laboratory space and 40–50 full-time research positions. That means that an active researcher (such as Francis Collins, a prominent University of Michigan geneticist who is rumoured to be the leading candidate) could transplant an entire laboratory. "The only place we're limited is in salary", Healy says.

Healy also favours an active recruitment campaign rather than one that simply places advertisements and waits for a response. "Happy, productive people don't read the help-wanted ads", she says. She has encouraged search committees to contact potential applicants and gauge their interest and she has halted the onerous practice of asking all prospects to fill out government application forms in triplicate. Now committees wait until they know who they want before proceeding with the paperwork. Although such changes may seem slight, they are intended to show that NIH can compete head-to-head with the outside world.

Other changes making NIH more

attractive include:

■ **Higher salaries.** Postdoctoral researchers can now make up to \$50,000, more than in academia. In addition, starting in January 1991, salaries for the government-wide Senior Executive Service (SES), for which NIH has 195 positions, rose by an average of 20 per cent, to \$144,000 for a researcher with a medical degree. By comparison, a medical school department chairman in the basic sciences earns an average base salary of \$133,000 (although many are also compensated for giving up clinical work), according to the latest survey by the American Association of Medical Colleges.

■ **More laboratory space.** Cramped conditions are a chronic problem, and many hallways are crowded with refrigerators and unused equipment. But things are looking up. Last week NIH broke ground for an administrative complex named after Representative William Natcher (Democrat, Kentucky), the chairman of one of the two committees in Congress that sets NIH's budget. Another new building, named after the late Representative Silvio Conte, is expected to open by the end of the year.

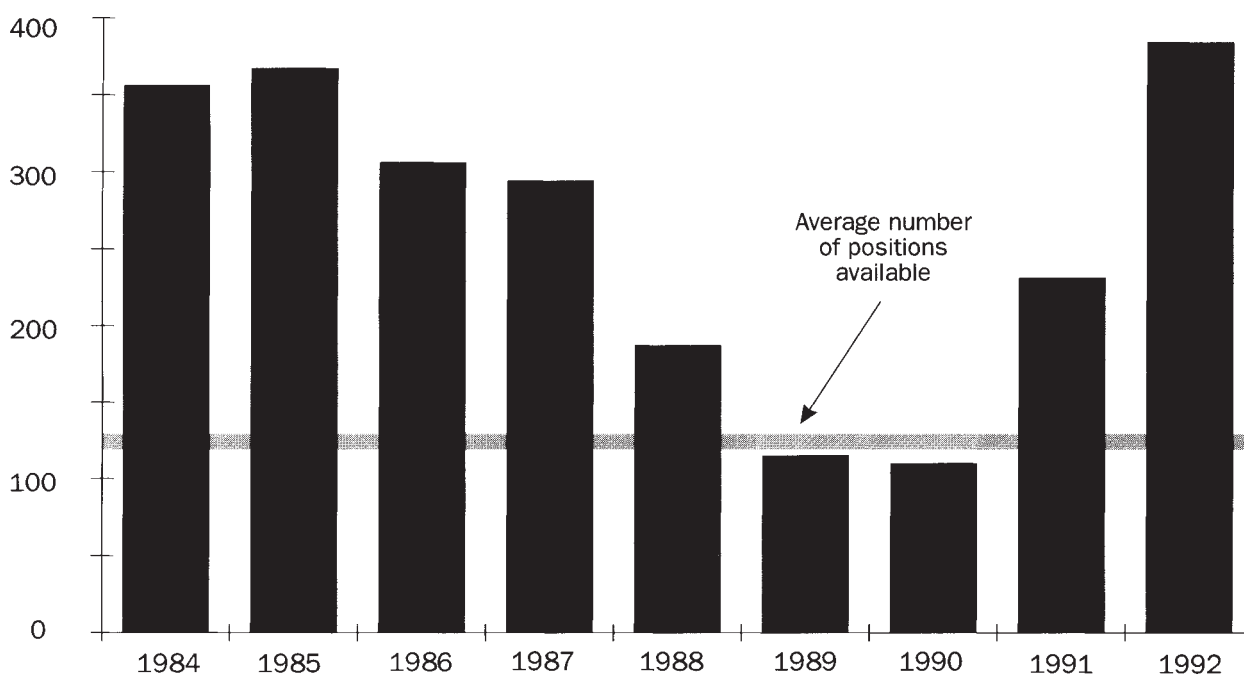
■ **Simplified consulting rules.** The new regulations, expected to take effect later this year, would set an annual limit of \$25,000 in income from one institution for any combi-



Lance Liotta: new intramural director aims to change perceptions.

nation of consulting and lecturing. At present, an NIH scientist can earn no more than \$12,500 for consulting and \$12,500 for lecturing. The new rules will reduce approval time and paperwork and eliminate the need for scientists to count half their consulting fees as lectures. Government regulations still require layers of approval, however; unlike university researchers, NIH scientists must work a full day on only approved topics to avoid running afoul of

Declining applications for MD postdoctoral jobs



NIH-wide applications for MD clinical research training positions fell so low during the late 1980s that many institute and division directors began to recruit on their own, bypassing the central office, from which these statistics come. In 1990, for example, the 92 applications to the central office are actually fewer than the number of positions that were available at the time.

the payment cap brings the NIH scientists closer in line with their colleagues at public universities.

■ A new Office of Education. The office, which opened in June 1990, offers programmes for postdoctoral researchers on the belief that such nurturing will entice them to stay. In the last year and a half, the number of accredited MD residencies NIH offers has doubled. The six new ones, in sub-specialties of internal medicine such as haematology and infectious diseases, have been the most popular and account for the majority of increased applications. With research institutions increasingly looking for sub-specialties experience, the new residencies should make researchers more marketable both outside and inside NIH, says Michael Fordis, the office's director. In addition, post-doctoral candidates can be as much as five years past their degree (the limit had been three years) and can remain for as long as seven years instead of three. The changes give NIH greater flexibility in hiring those who are changing careers or returning to work.

■ Smarter recruiting. Rather than advertising postdoctoral positions individually, the new education office now takes out full-page advertisements listing dozens of positions throughout NIH and highlighting opportunities in specific areas that crisscross several institutes. It has also created an electronic database of NIH positions and prints instructions for accessing the database by modem in its advertisements. Such salesmanship has sharply increased the number of inquiries about postdoctoral positions and resulted in 946 applications for 67 positions — a 14:1 ratio that is one of the highest in NIH history.

■ Earlier tenure tracks. NIH officials are now making a point of identifying the young researchers most likely to be offered tenure within a year or two and giving them extra laboratory resources and responsibilities. As Fordis puts it: "Young people going into research need a sense that they're going to be good at it." So rather than waiting the usual three or four years to select tenure-track candidates — by which time many young researchers at NIH are already looking elsewhere — "we want to show them early on that there's a commitment to their career".

Officials are banking on NIH's reputation for quality and the opportunity it offers for both basic and clinical research. Half the clinical research beds in the United States are at NIH, and collaboration across disciplines often requires only a walk down the hall instead of a trip cross-country.

NIH is also counting on the fact that times are tough and that its disadvantages are minor compared with such chronic outside worries as rejected grant applications and university cutbacks. Twenty-five years after the Vietnam draft, NIH once again looks like a safe refuge.

Christopher Anderson

NIH looks inward and outside for ways to help minorities

The dual nature of the National Institutes of Health as both funding agency and in-house laboratory is well-illustrated by its efforts to increase the number of scientists from minority groups. At the same time that NIH wants to give US institutions millions of dollars to encourage minority students to go into biomedical research, its intramural programme is struggling to improve a



John Ruffin

sorry record of hiring and promoting minority scientists on the Bethesda, Maryland, campus. Congress seems favourably disposed to give NIH \$20 million in the fiscal year starting next month for new training programmes as part of a \$45-million initiative to improve the health of minorities. One aim is to attract minorities into graduate programmes by linking historically black colleges with major research universities. The students would simultaneously enrol in a master's programme at the black college and a doctoral programme at a nearby research university. Other initiatives would create a 'bridge' for students at two-year community colleges to continue their education at four-year undergraduate universities and would support high school and middle school students in programmes intended to whet their appetites for more science as undergraduates.

At the same time, NIH would continue its support for two programmes begun in the 1970s, the Minority Access to Research Careers and Minority Biomedical Research Support, as well as a rapidly growing summer programme, a supplemental grants programme that adds minority scientists to existing research projects and in-service training for high school and undergraduate minority faculty. The new projects will be run from the two-year-old Office of Minority Affairs, headed by John Ruffin, which has asked for an additional \$9.8 million in 1993.

Although these programmes will help the overall research community, they will do little to improve the lot of the tiny number of minority scientists in the \$1-billion intramural research programme. NIH officials are reviewing a report submitted last month by outside consultants that paints a bleak picture of the chances of minorities becoming tenured scientists in the staff fellows programme, which is a primary way of

attracting young talent to NIH. Although the odds in general are long — only about one in ten fellows win a permanent position at NIH by the end of their seven-year probationary period — the number of minority scientists who have done so can be counted on one hand.

The report, which took two years to complete, was carried out for NIH's Office of Equal Opportunity, which is responsible for seeing that all employees are treated fairly. Independently, the scientific directors of the intramural programmes at each NIH institute decided earlier this year to look at the issue of recruitment, retention and promotion of minority scientists and formed a committee led by Bruce Chabner of the National Cancer Institute. That committee is expected to use the consultants' findings in its own analysis, due by the end of the year.



Bruce Chabner

One problem that institutions face is the stiff competition for the small number of qualified minority scientists, while students must overcome the high cost of earning an advanced degree, both in tuition fees and in a temporary loss of earnings. One solution, says Chabner, is to support those who choose careers in research through tuition subsidies or by forgiving student loans.

But improving the lot of those who join NIH is equally important. Chabner estimates that there are no more than a dozen tenured minority scientists in the intramural programme, and some met each other for the first time as members of Chabner's committee, according to Michele Evans of the National Institute on Aging, who spent two and a half years as special assistant for minority affairs in the cancer institute.

There is no question that better answers are needed, says Bernadine Healy, director of NIH. Admitting that she was "disappointed" by what the consultants reported, Healy says that she has asked her staff to think about reasonable targets and "to become more aggressive" in recruiting, retaining and promoting minorities. "We have a problem", she says, "and the institute directors understand that this is one thing that they will be judged on" when their job performance is reviewed.

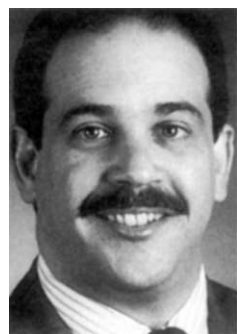
Jeffrey Mervis

Transfer of technology is booming business as NIH asks companies to help themselves

Washington. A telephone call to the office of technology transfer (OTT) at the National Institutes of Health gives a clue to the changes going on inside. Whereas three years ago such a call would be answered by one of the office's three employees, it is now picked up by a voice-mail system worthy of a large corporation. Inquiries are electronically directed to one of four divisions with an over-

all staff approaching 50. Half a decade after Congress first allowed government agencies to commercialize their science, technology transfer has become big business at NIH.

A long history of collaboration with pharmaceutical companies has allowed NIH



Reid Adler

to move into technology transfer more quickly than most other federally funded laboratories. Since its creation in 1987, the NIH office has broadened its responsibilities to cover technology transfer programmes at two other science agencies within the US Public Health Service (PHS) — the Alcohol, Drug Abuse and Mental Health Administration and the Centers for Disease Control.

Government technology transfer has its roots in the late 1970s, when Congress became interested in using the US patent system to promote the commercialization of inventions arising from federally funded research. In 1980 Congress changed the patent law to permit some such transfers. Then, in 1986, it broadened the legislation with the Federal Technology Transfer Act, which authorizes federally funded laboratories to enter into Cooperative Research and Development Agreements (CRADAs) with outside organizations, primarily industry, and to patent and licence their inventions.

Obtaining patent protection for government-made inventions lies at the heart of NIH's technology transfer activities and is strongly supported by Bernadine Healy, the director of NIH. It provides for public disclosure of inventions and protects the rights of institutes within NIH and of individual inventors, while stimulating interest in the commercialization of government-supported inventions by pharmaceutical and biotechnology companies. However, NIH created a furore last summer when it took the unprecedented step of filing patent applications on thousands of partial gene sequences (complementary DNA sequences) of

unknown function.

Although CRADAs provide NIH scientists with additional resources, researchers from outside organizations are expected to make a significant intellectual contribution in jointly developing the technology. Government researchers may also receive a share of the royalties on any products licensed as a result of the collaboration.

Marc Schneebaum, vice-president and chief financial officer of Genetic Therapy Inc. (GTI), says that companies were unwilling before 1986 to collaborate with federally funded laboratories because "they couldn't protect their proprietary position". With the introduction of CRADAs, outside organizations can now secure exclusive licence agreements on any inventions that result from CRADAs. Although NIH researchers in a CRADA retain the right to publish their results, publication can be delayed to stake out a corporate claim to patent rights.

In the three and a half years since Reid Adler took up the post as director of OTT, the staff has grown in size from 2 to 42. But in the past six months, the number of CRADA applications has risen by 50 per cent to 7–10 submissions a month. Adler believes that the office needs to grow further, to at least 60 people, to handle the patenting, licensing and cooperative research programmes, to stay in touch with institutes within NIH and to keep industry abreast of possible collaborations.

Organizational changes are under way to simplify the technology transfer process. Since May, NIH's patent group, formerly part of the Office of the General Counsel is has joined OTT. Adler plans to establish joint patent and licensing teams to manage the portfolio of inventions and expand the services available to each institute. Although the number of patent application filings has remained level during the past three years, there are about 1,200 active patent applications at OTT. As the office files only 200 or so a year, half of the applications were filed before Adler arrived at NIH. The patent team plans to thin out the portfolio during the next six months and to be more selective in filing future patent applications.

What is in it for industry? For smaller companies, Adler says that CRADAs are a cost-effective way to increase the value of a company's research and development. For example, when HealthCare Investment Corporation started MedImmune Inc., it located it in Gaithersburg, Maryland, near the NIH campus, with the hope that the proximity would spur collaborations that could lead to new products. MedImmune now has seven CRADAs with NIH, five in immunotherapy and vaccine development. CRADAs have

been "extremely useful to us", says James Young, vice-president for research and development at MedImmune, allowing "us to expand our resource base" to include the entire NIH intramural science programme.

For GTI, also located in Maryland, NIH and its 540-bed hospital serve as the company's 'clinical outlet'. The company was co-founded in 1986 by W. French Anderson, one of the pioneers of gene therapy and formerly of NIH's National Heart, Lung and Blood Institute (NHLBI). GTI signed its first CRADA with NIH in 1988 and now has five collaborations to test gene-therapy approaches to a wide range of diseases, including adenosine deaminase (ADA) deficiency, AIDS, blood disorders, cancer and cystic fibrosis. In 1990, GTI became the first company to receive a licence from NIH for technology developed under a CRADA carried out in conjunction with NHLBI and the National Cancer Institute.

But some other companies are still sceptical that government technology can be easily made into commercial products. Mark Furth, vice president of Regeneron Inc. in Tarrytown, New York, says that "dealing with the government is too slow to rely on" and that the company tries to "let science drive what we do" rather than promising to work with a government agency. Regeneron is developing compounds for the treatment of various neurological disorders such as Alzheimer's disease, amyotrophic lateral sclerosis (ALS) and Parkinson's disease. It has more than 100 collaborations with universities worldwide, but only one CRADA (with a protein structure group at NIH's National Institute of Diabetes and Digestive and Kidney Diseases).

Most discussions concerning CRADAs involve the scope of the research and what each party will contribute, but the issue of pricing is also being closely watched by industry. An existing clause allows NIH to exercise some control over the pricing of products developed through the programme, and industry feels that it is a disincentive to becoming involved in CRADAs.

For the past decade, federally funded laboratories have grappled with how to create, run and organize technology transfer programmes. What is needed in the 1990s, Adler says, is greater professionalism among technology transfer staff, a better allocation of resources and increased collaboration between agencies within PHS. He hopes that a new association of federal technology transfer executives (see *Nature* 358, 362; 1992) will promote the sharing of ideas and information and eliminate the need to "reinvent the wheel" at every step of the technology transfer process.

Diane Gershon

Gene therapy research is spread across many disciplines and institutes at NIH

In 1990, W. French Anderson, R. Michael Blaese and Steven Rosenberg of the US National Institutes of Health began the first clinical test of a human gene therapy. Now, only two years later, a host of other NIH intramural scientists are working on gene therapies for diseases they have studied for years. And several are nearly ready to submit protocols to NIH's Recombinant DNA Committee (RAC), which judges the ethical

and scientific soundness of proposed trials for gene therapy and gene transfer.

The process of winning RAC approval for that first gene therapy and, in 1988, for the first gene transfer has helped to shape gene therapy research across the NIH campus. Scientists have learned what data RAC demands and what ethical and scientific questions it might raise. Although many NIH researchers had worked on gene therapy before 1988, approval of the first protocol contributed to its spread across institute lines.

Michael Blaese

Although gene therapy was first seen primarily as a treatment for genetic diseases, those working on nongenetic diseases now believe that it may be one of the best ways to treat many illnesses. "Although it might be quite a while before it's actually put into practice", says David Dichek of NHLBI, who is working on gene therapies for cardiovascular disease, "I thought it had promise to be superior to existing therapies."

Widespread collaboration among intramural and extramural scientists has fostered gene therapy research in both settings (see accompanying story). "It took the back-

grounds of many different people to make the whole story come together", says Blaese. "That's the fun part of doing research at NIH." In the list that follows, credit for each project must be shared with dozens of other scientists whose names are not mentioned.

Following are some of the gene therapies being developed at NIH:

■ Cancer

In the first trial of a gene therapy for cancer, Steven Rosenberg at the National Cancer Institute (NCI) is treating seven melanoma patients with tumour-infiltrating lymphocytes (TILs) carrying the gene for tumour necrosis factor (TNF), a protein believed to restrict a tumour's blood supply. Rosenberg is treating five more melanoma patients with tumour cells containing the gene for either TNF or interleukin-2, both of which help to identify cancer cells as foreign. He recently received permission to test

the tumour-cell therapy on patients with renal, colon and breast cancers.

R. Michael Blaese, also at NCI, is working on a therapy based on the drug ganciclovir, normally used to treat herpes and an infection, cytomegalovirus, associated with



Arthur Nienhuis

AIDS. The viral enzyme thymidine kinase switches on ganciclovir, killing the virus, while human cells escape unscathed because they lack the viral kinase. Blaese is developing techniques to dose tumours with the viral kinase gene, so that the gene-carrying tumour cells would manufacture the activating enzyme and become susceptible to ganciclovir. The viral thymidine kinase therapy has already been approved by the RAC, and several similar therapies are under development.

Arthur Nienhuis of NHLBI is investigating the use of gene therapy to protect bone-marrow and peripheral blood stem cells from anticancer drugs. Nienhuis has found that mouse stem cells altered to carry the multi-drug resistance gene survive chemotherapy better than the unaltered stem cells. In humans, this therapy may allow doctors to stamp out stubborn tumours without killing the vital — and vulnerable —

stem cells. Nienhuis has started trials with monkeys and hopes to gain approval for human trials within six months.

■ AIDS

Robert Gallo of NCI is developing several gene therapies that will attack the regulatory machinery of the human immunodeficiency virus (HIV). In one study, lymphocytes are engineered to contain 25 to 50 copies of the TAR (tat activation responsive ele-



Richard Morgan, top, and Robert Gallo, left.



ment) gene, a crucial element for switching on viral replication. When HIV invades an immune cell and tries to reproduce, confusion between the viral TAR and the cell's multiple copies of TAR effectively blocks replication.

Richard Morgan of NHLBI is working on a gene therapy to prevent HIV from attaching to and eventually destroying T-cell lymphocytes. Morgan is designing a T-cell that produces CD-4, the T-cell antigen that binds HIV, and dumps it into the bloodstream. There the molecules would act as decoys to prevent HIV from ever binding with the CD-4 on the surface of the T cells. Morgan plans to submit a protocol to RAC by the end of the year.

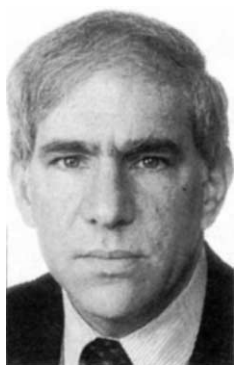
■ Hereditary disorders

The first recipient of gene therapy was a 4-year-old girl lacking adenosine deaminase (ADA), the immune cell enzyme whose absence leads to severe immunodeficiency. The girl received the first T cells carrying the ADA gene in late 1990; now she and an 11-year-old girl whose treatment began in early 1991 have much stronger immune systems, as indicated by their production of antibodies against influenza virus administered in vaccinations.

Meanwhile, Nienhuis and Blaese have adapted the ADA treatment to peripheral blood stem cells and hope to begin clinical



Steven Rosenberg



Ron Crystal

tests on the two girls by the end of the year. If ADA-deficient patients could receive progenitor stem cells containing the ADA gene instead of the much shorter-lived T-cells, they would need fewer transfusions — and eventually none.

A gene therapy for cystic fibrosis is being tested in primates by Ronald Crystal of NHLBI. Cystic fibrosis results from a mutation affecting the epithelial cells lining the airways of the lung. Crystal realized that virus containing the desired gene would need to be applied directly to airway cells, so he decided to introduce the virus through a bronchoscope rather than with injections of virus-carrying cells, a more widely used technique. Crystal hopes to submit a protocol to the RAC within six months.

Patients with chronic granulomatous disease produce no superoxide anion, which is



Top to bottom:
Harry Malech,
Ed Ginns, and
Stefan Karlsson.

used by phagocytes to kill invading microorganisms. As a result, patients suffer frequent life-threatening fungal and bacterial infections. Harry Malech of the National Institute of Allergy and Infectious Diseases is working on inserting the genes involved in superoxide production into peripheral blood progenitor cells.

Two NIH researchers are working on gene therapy for Gaucher's disease, a hereditary lack of the enzyme used by lysosomes to break down the common lipid breakdown product glucocerebroside. Patients suffer from weak bones, anaemia and, in some cases, nerve damage. Stefan Karlsson of the National Institute on Neurological

Disorders and Stroke is developing bone marrow stem cells that would carry the enzyme and generate normal lysosomes. Edward Ginns of the National Institute of Mental Health is working on a two-part strategy: first, clear patients' bodies of accumulated fat with a recently introduced drug for Gaucher's, then keep lipid levels down by injecting patients with bioengineered non-stem cells such as fibroblasts or endothelial cells. Both Karlsson and Ginns hope to submit proto-cols to the RAC soon.



Robert Nussenblatt

to high levels of ornithine in the blood. Nussenblatt's research team is working on inserting the enzyme into lymphocytes that would collect and break down the excess ornithine in the bloodstream.

Morgan is also experimenting with a therapy for haemophilia. He is inserting the gene for clotting factors XIII and IX into different cell types *in vitro* to learn whether the genes will retain their ability to synthesize protein.

■ Cardiovascular disease

Clots often form in synthetic blood vessels used for bypasses, so Dichek of NHLBI is trying to create a bioengineered endothelial cell that will stick to the vessel walls and produce clot-dissolving protein. Dichek also hopes to use bioengineered viruses to prevent the reclogging of the arteries that commonly follows angioplasty. Through studies on animals, he has learned that catheters can be used to deliver gene-laden carrier viruses to the cells that make up artery walls. Dichek hopes to increase the amount of virus entering the cells so that the technique could be used to treat patients with genes that would inhibit clogging.



David Dichek

Traci Watson

Flourishing genes

Gene therapy research is flourishing not only on the campus of the National Institutes of Health in Bethesda, Maryland, but also at research centres across the country. This month, for example, the National Cancer Institute will award the first grants in a four-year, \$20-million gene therapy programme.

The following institutions have scientists working on gene therapy or gene transfer protocols that have been accepted by NIH's Recombinant DNA Committee:

Baylor College of Medicine, Houston;
University of California, Los Angeles;
Indiana University at Indianapolis;
M.D. Anderson Cancer Center, Houston;
Memorial Sloan-Kettering Cancer Center, New York;
University of Michigan at Ann Arbor;
University of Pittsburgh (Pennsylvania) School of Medicine;
St. Jude Children's Research Hospital, Memphis, Tennessee;
University of Washington, Seattle.

NIH to help women return to research

Washington. Female biomedical researchers who leave science to raise children or attend to other family needs can get help in resuming their careers through a new \$1 million programme sponsored by the US National Institutes of Health (NIH). NIH officials hope the programme, the first to encourage women to return to science, will increase the number of female scientists in top-level research positions.

The programme, sponsored jointly by the Office of Research on Women's Health and the Office of Extramural Research, will award 10 to 15 women as much as \$40,000 annually for one year to do research with scientists already holding NIH grants. The project will also pay up to \$10,000 for supplies and travel for each woman, whose work must be related to the project and serve to "refresh existing research skills and knowledge." Candidates must hold a doctoral degree and have completed most or all of their postdoctoral training.

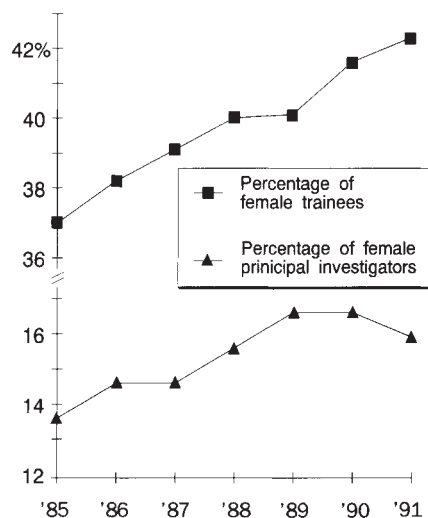
Funding requests must come from the researchers with whom the women want to work. This year, 55 applications have been filed, and NIH hopes to hand out the money before the end of the current fiscal year on 30 September.

The programme is intended to rectify a disturbing statistic: although women hold more than 40 per cent of NIH's biomedical training positions, only 19 per cent of the

pool of principal investigators of NIH research projects are women. It is not known why so many women leave research. By aiming a programme at women who left science for family reasons, NIH hopes to test the theory that child-raising accounts for much of the attrition. The women's health office has already promised to fund the programme again in 1993 and, if its budget permits, in 1994.

Although NIH is praised by women for addressing an important problem, there are

Number of female principal investigators still lags



The percentage of women holding NIH trainee appointments still far exceeds the percentage of women receiving NIH research project grants. Both numbers have increased over the last 10 years, but the number of female principal investigators have failed to keep pace with the number of trainees.

questions about how well the programme will work. Time is short: the official announcement appeared only one month before the applications were due, and it has attracted little public notice. And the process of finding applicants is difficult. Walter Schaffer, an officer of the extramural research office, admits that "there isn't any really good way" for them to do so.

The most serious obstacle may be beyond NIH's control: will scientists be willing to take on women who, for five or ten years, have not read the literature and have had little contact with other researchers? "There is a real prejudice" towards people who lack the right training, says Catherine Didion, executive director of the Association for Women in Science in Washington DC. If so, NIH may find that the programme — at least in its current form — is not a true test of the existence of a pool of women hoping to return to science. **Traci Watson**

Hughes woos future MDs by offering space at bench

If widespread concern were all that mattered, the shortage of physicians entering biomedical research that has existed for the past 20 years would have been remedied long ago. But it has persisted despite several programmes by the US National Institutes of Health (NIH) to encourage MDs to undertake research.

Several years ago, the private sector decided to lend a hand. One such effort operates on the NIH campus itself: the Research Scholars Program, run by the Howard Hughes Medical Institute (HHMI). The institute, near NIH in Bethesda, Maryland, offers medical students a one-year sabbatical to do research in an NIH laboratory in the hope that students with first-hand experience in the laboratory will be more likely to carry out research after they graduate.

The programme, now seven years old, has placed 230 students from 73 medical schools in various NIH laboratories. It has proved so popular that HHMI has slowly expanded the class size from 23 to 50, and some students, absorbed in their research, elect to stay a second year. Under a programme begun in 1989, HHMI also awards one-year grants for medical students to work at an academic or nonprofit research centre anywhere in the United States.

Created from a gift of Hughes Aircraft stock from its founder that has grown in value to over \$6 billion, HHMI supports the programme out of a conviction that physicians have an invaluable role in studying human health and disease. Hughes is also motivated by statistics showing that the percentage of MDs among applicants for research grants from NIH has dropped from 27 per cent in 1980 to 23 per cent in 1991. Only 2.2 per cent of doctors spend most of their time in research, according to the American Medical Association.

One obstacle is the difference in the salaries of researchers and physicians — an important issue for a medical student graduating with debts that average \$50,000. Another deterrent is the long training required to do research: medical training alone demands seven to eleven years, and many hesitate to postpone their careers by tacking on a few years of postdoctoral

work. The Hughes programme addresses the problem of ignorance: medical schools make little provision for students to carry out research, and without such experience students lack sufficient information to choose between the laboratory and the bedside.

Students can put their medical background to the test in one of almost 300 NIH laboratories under the guidance of a tenured investigator. Some participants start and



The Cloisters on the NIH campus.

complete their own projects during their stay, while others take over experiments already in progress. Students spend 40 to 80 hours per week in the laboratory and must give a presentation of their work to their fellow students. Participants live on the NIH campus in a beautiful former convent built in the 1920s by the Sisters of the Visitation.

It is too early to judge the impact of the programme. Some of the members of the first class are still residents; of those who have finished, approximately half are doing postdoctoral research, including three who returned to NIH.

Whatever their motives, all appear to benefit from their year at NIH and to return to medical school with more knowledge about the causes of disease and the mechanisms of drugs. Even if the programme fails to swell the ranks of clinical investigators, few NIH officials would quarrel with the value of that outcome. **Traci Watson**

For more information about these programmes, contact Lainey Wittenberg, HHMI, One Cloister Court, Bethesda MD 20814-1460.

Who needs a new magnetometer?

To every measurable effect there seems to be an instrument whose purpose sometimes seems to be simply to measure the measurable causes of those effects. The purpose of a neat new magnetometer is not yet clear.

THERE seems to be a general principle in the physical sciences that each newly discovered effect will eventually spawn an instrument of some kind. Sooner or later, somebody will make a microscope of each new particle of matter discovered, for example. More generally, the principle seems to be no more than the truism that if a cause has an effect that can be measured, the effect can be used to measure the cause. That is the spirit in which a French group has now built a magnetometer based on the interaction between a magnetic field (such as the Earth's) and the behaviour of a laser light.

As described (in *Phys.Rev.Lett.* **69**, 909–912; 3 August 1992), it has been thoroughly Cartesian. Those concerned are Fabien Bretenaker, Bruno Lépine, Jean-Charles Cotteverte and Albert Le Floch, all at the CNRS unit for quantum electronics at the University of Rennes.

The Cartesian point of the argument amounts to a taxonomy of magnetometers. There are classical magnetometers and quantum magnetometers, of which the latter are capable of inherent micro-sensitivity. But quantum magnetometers are themselves of two kinds. There are those based on the properties of individual atoms, as in the use of the precession frequency of, say, protons to measure the magnetic field to which they are externally exposed; the strength of the signal is bound to be an average of the values generated by many differently placed atoms, so that inhomogeneities of the field are untidily smeared out. And then there are microscopic quantum detectors best typified by the SQUID, in which the quantized collective motion of pairs of electrons in a superconducting element provides sensitivity almost free from noise. But lasers are also collective quantum systems; could they not do as well as SQUIDS? It would be interesting to know what set Bretenaker and his colleagues on the path to their magnetometer, or whether they simply had a hunch.

In practice, the laser they use is a curious animal, consisting of a 3 mm hole drilled along the axis of a cylindrical block of a proprietary material called Zerodur which is nonmagnetic, but is pleasantly stable to thermal change. The result is that the length of the 60 cm laser cavity remains

essentially constant. A further curiosity is that the ends of the cavity are optically closed by what are called quasi-mirrors — pieces of glass that would have been mirrors had they been silvered, but which in reality reflect 3 per cent of the light incident on them. The cavity is filled with a mixture of helium-3 and neon-20 and is activated by an electrical discharge. The active principle is the neon-20, which has optical transition in the infrared at 3.9 μm .

The sense in which such a system is a macroscopic quantum system is crucial. "Laser", after all, means "light amplification by stimulated emission of radiation". During a single passage of a matching photon along the narrow cavity, the excited neon atoms are stimulated to emit no fewer than 700 further photons. Evidently, it is of little concern that 97 per cent of them are lost at each end of the 60 cm track. The quantum system that matters in this case is not the excited neon atoms, but the electromagnetic wave that propagates to and fro in the narrow tube, dimensions of which have been fixed to allow only the simplest possible wave with symmetry along the axis.

How can such a system be used as a magnetometer? An external magnetic field will bring about a splitting of the energy levels of the neon-20 atoms, but what will happen to the electromagnetic field or laser beam? The Faraday effect, of course.

This is indeed the standard means of measuring the strength of interstellar magnetic fields, by the degree to which it causes the rotation of the plane of polarization of a microwave signal from a distant quasar or other source.

With the Rennes laser, what happens is much the same. Given its symmetry, it is sensitive only to the components of magnetic fields along its axis. Since there is no reason why original laser should be polarized, it can be thought of as what it is — a mixture of oppositely polarized circular waves. But then the effect of the axial magnetic field, which would split the neon energy levels through the Zeeman effect, is to change the frequencies of the oppositely circularly polarized waves, if only slightly. And then it is possible to measure the differences between the two frequencies by beating the two waves against each

other. The result is that a magnetic field comparable with that of the Earth shows up as a frequency difference measured in kilohertz. With the laser built at Rennes, 200 milligauss shows up as just under 250 kilohertz.

That, then, is the essence of the magnetometer. Its virtue is that it substitutes for the measurement of magnetic field the much more sensitive measurement of the frequency difference of two parallel beams of infrared radiation, whose mutual interference can be followed by passing them through a polarizing device and a detector. The response of the instrument appears exquisitely a linear function of the magnetic field, at least for field strengths of the order of 200 milligauss. The only snag is that the laser, in the absence of an external field, appears to lock into a preferred mode from which it must be dragged by the Faraday effect of a field of at least a few milligauss. The authors guess that the preferred locking position is determined by geometric irregularities in the quasi-mirrors.

So who will use the new magnetometer? The way things are, this would not be the first time that a spanking new instrument with exquisite sensitivity had found no uses. What Bretenaker and his colleagues find is that their device is quickly responsive to small changes in the magnetic field — 100 microseconds seems to be enough. And dynamic measurements, in which an artificially induced field of known frequency is added to the field to be measured allows for a sensitivity for a sensitivity of 10^{-5} gauss.

It should not take very long for interested people to think of plausible uses for this new instrument. The authors themselves suggest that investigations of the supposedly quickly changing magnetic fields in advance of earthquakes are a natural starting point. Presumably, they could study the natural variation of the Earth's magnetic field, with its ionospheric solar-terrestrial antecedents. In each case, the volumes of data generated would be huge — up to 1,000 measurements a second.

The era of multi-micro-magnetism seems at hand. And somebody is bound to have a laser magnetometer in a spacecraft very soon.

John Maddox

Superfluid gyroscope

George B. Hess

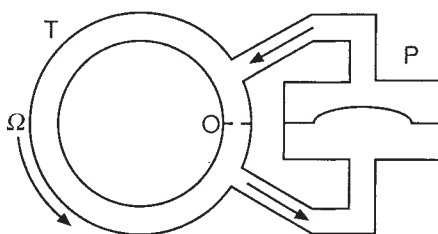
THE quantum properties of liquid helium may provide a precise technique for monitoring the rotation of the Earth. This possibility arises because circulation of superfluid helium is quantized, so that a torus filled with liquid helium can give a stable reference state of zero rotation with respect to inertial space. A superfluid-helium gyroscope (SHEG) based on this principle has been suggested in the past¹, but no suitable technique was available for measuring the motion of the toroidal container relative to the helium inside. The critical elements required for this measurement have been demonstrated in experiments by Oliver Avenel and Eric Varoquaux^{2,3}. In the light of this work, Richard Packard and Stefano Vitale now review⁴ the design parameters and potential sensitivity of a SHEG, concluding that, unless noise sources not yet understood intervene, the sensitivity should be sufficient for interesting geodesic observations.

The 'circulation' around a thin torus is the flow velocity times the length of the path along the centre of the toroidal channel. For superfluid ⁴He, this quantity is constrained by phase coherence of the superfluid state to have a value $n\kappa$, where n can be any integer and $\kappa = h/m = 1.0 \times 10^{-7} \text{ m}^2 \text{ s}^{-1}$ (h is Planck's constant, and m is the mass of a helium atom). If the toroidal container is rotating about its symmetry axis, owing for example to the rotation of the Earth, but $n = 0$, then the container is moving relative to the helium. To an observer in the rotating laboratory, the superfluid appears to be flowing around the torus. Even if n takes the integer value which allows the helium to approximate most closely the rotation rate of the container, there will in general be a residual relative velocity, which a clever observer can hope to measure. (It would then be simple to determine n by reorienting the torus in the rotating laboratory.)

The first step in making the measurement is to constrict the flow channel at one section. The velocity in the constriction will increase relative to that in the rest of the torus in inverse proportion to the cross-sectional areas. By going to the extreme case of a submicrometre orifice in a thin partition, it is possible to increase the velocity a millionfold.

The measurement of this increased velocity is carried out by adding to the apparatus a diaphragm pump (see figure), which superimposes an alternating flow on the rotation-induced flow. An extremely sensitive sensor of the displacement of the diaphragm allows the

magnitude of the alternating flow to be monitored. Frictionless flow of superfluid ⁴He persists only up to a certain critical velocity, which for a small orifice is typically a few metres per second. When this velocity is reached, the orifice generates a quantized vortex ring or loop, which sweeps across the flow path, extracting energy. This so-called phase-slip event changes the circulation quantum number n by 1. Flow in the orifice



Outline of the superfluid helium gyroscope. The main components are: the rotation sensing torus, T; partition containing a small hole, O; diaphragm pump, P, with diaphragm displacement sensor not shown. The entire apparatus is filled with liquid helium at a temperature below 1 K.

owing to rotation changes the pump flow needed to reach the critical velocity, and this forms the basis for measuring the relative rotation. This is a close analogue of a superconducting device called the radiofrequency SQUID (superconducting quantum interference device), which sensitively measures magnetic flux.

Avenel and Varoquaux first demonstrated that single-quantum phase slip events can be detected², and then showed that conditions can be established in which only single-quantum events occur³. The SHEG output, which is the shift in the peak alternating velocity, is proportional to the angular velocity of the apparatus, modulo $\kappa/2\pi R^2$ (R is the path radius). Thus it is a sawtooth function of angular velocity Ω . This modulation is washed out if multiple-quantum events occur. The sensitivity can be improved by increasing the torus's radius and using a multiple-turn coil. Packard and Vitale contemplate a ⁴He SHEG of 1 metre diameter with 32 turns; little is gained by adding more turns, as the circulation becomes less concentrated at the orifice.

A more favourable situation exists if one uses superfluid ³He-B instead of ⁴He, because the phase slip process for this superfluid is a continuous, non-dissipative process, analogous to that in a Josephson junction between superconductors, with a low critical velocity, and the possibility of multiple-quantum events can be eliminated. Avenel and

Varoquaux³ demonstrated this, too. In this case the apparatus must be operated at a temperature below 1 millikelvin.

Varoquaux *et al.*⁵ projected a resolution for a ³He version of the SHEG of order $10^{-6} \times \Omega_E$, where Ω_E is the angular velocity of the Earth's rotation. Packard and Vitale estimate a value five times smaller, with a further large reduction possible if the sampling frequency can be high and the intrinsic noise in the phase-slip process is amenable to averaging.

The complexity of the required cryogenic system and vibration isolation is likely to limit a SHEG to laboratory or fixed station use for the foreseeable future. If a sensitivity of order $10^{-9} \times \Omega_E$ were attained, one could measure fluctuations in the Earth's rotation rate. The Earth's rotation is currently monitored by external observations: classical astrometry, very-long-baseline radio interferometry, lunar or satellite laser ranging, or satellite doppler tracking (such as the Global Positioning System)⁶. These provide a resolution of order $10^{-9} \times \Omega_E$ from a few days of observations (less for VLBI) and suitable data reduction. General relativity predicts a 'gravitomagnetic' difference between the observationally and inertially measured rotation rates (the Lense-Thirring effect) of about $4 \times 10^{-10} \times \Omega_E$, but study of this effect requires absolute calibration exceeding this accuracy⁷.

In any application, the SHEG would have to compete with more conventional gyroscopes. Electrostatically supported rotating-mass gyros, developed for inertial navigation, may be able to resolve 10^{-5} or $10^{-6} \times \Omega_E$ on a timescale of one hour. Much greater precision is possible in free-fall: the electrostatic gyros of Gravity Probe B are expected⁸ to have sufficiently low drift in a drag-free satellite to allow measurement of the Lense-Thirring effect to a few per cent in one year.

Fibre-optic interferometer gyros and commercial ring-laser gyros have drift rates of 10^{-2} to $10^{-4} \times \Omega_E$ on a timescale of one hour. These devices are based on the Sagnac effect, the property that light beams travelling in the two senses around a circular or polygonal path take

1. Cerdonio, M. & Vitale, S. *Phys. Rev.* **B29**, 481–483 (1984).
2. Avenel, O. & Varoquaux, E. *Phys. Rev. Lett.* **55**, 2704–2707 (1985).
3. Avenel, O. & Varoquaux, E. *Phys. Rev. Lett.* **60**, 416–419 (1988).
4. Packard, R. E. & Vitale, S. *Phys. Rev.* **B46**, 3540–3549 (1992).
5. Varoquaux, E. *et al. Physica* **B178**, 309–317 (1992).
6. Moritz, H. & Mueller, I. I. *Earth Rotation: Theory and Observation* (Ungar, New York, 1987).
7. Bonaldi, M. *et al.* in *Proc. 4th Marcel Grossman Meeting on General Relativity* (ed. R. Ruffini) 1309–1318 (North-Holland, Amsterdam, 1986).
8. Everitt, C. W. F. in *Near Zero: New Frontiers of Physics* (eds Fairbank, J. D. *et al.*) 587–639 (Freeman, New York, 1988).
9. Chow, W. W. *et al. Rev. mod. Phys.* **57**, 61–104 (1985).
10. Höling, B. *et al. Appl. Phys.* **B55**, 46–50 (1992).

different times to complete the circuit if apparatus is rotating. (Light travelling in the same sense as the rotation must go farther.) In a fibre-optic gyro, a coil of single-mode optical fibre, the two beams come from a single external laser, and are recombined to give interference fringes, which are shifted by rotation. If the laser itself is in the loop, then one has a ring laser, in which oppositely propagating modes have frequencies different by an amount proportional to the angular velocity⁹. Höling *et al.*¹⁰ have just proposed a ring laser of much increased area and power which might achieve resolution of geodesic interest.

A limited analogy can be drawn be-

tween the SHEG and the fibre-optic gyro, in which the rotation-induced circulation quantum number of the plays a role similar to the integer fringe shift of the latter. These are proportional, respectively, to $1/k$ and $2/c\lambda$, where c and λ are the speed and wavelength of light. The ratio of these quantities is about 10^9 . So the small magnitude of the quantum of circulation makes the SHEG attractive. The resolution attainable with each device depends ultimately on its noise and drift characteristics. □

George B. Hess is in the Department of Physics, University of Virginia, Charlottesville, Virginia 22901, USA.

CELL BIOLOGY

New cytoskeletal liaisons

Lawrence S. B. Goldstein and Ron D. Vale

BIOLOGISTS are accustomed to the idea that microtubules and actin microfilaments have different cellular functions and constituents, but over the past year it has become evident that the components and functions of the two are more intertwined than was once thought. This overlap is extended in reports by Lees-Miller, Helfman and Schroer¹, and by Clark and Meyer², on pages 244 and 246 of this issue. The authors have independently identified a distant relative of actin and show, respectively, that it is a component of the dynactin complex, which activates dynein-mediated organelle transport³, and of the centrosome, the microtubule-organizing centre of most eukaryotic cells.

These results are striking in view of recent indications that microtubules and microfilaments may serve similar or even partially redundant functions in some circumstances. Previous genetic and physiological studies indicated that animal cells use microtubules for transporting membrane organelles, whereas fungal, plant and protozoan systems use actin microfilaments for these tasks. This phylogenetic division broke down over the past year as it became clear that all eukaryotic cells may use both cytoskeletal systems for membrane transport. In extracts of the squid giant axon, where most organelle transport appeared to be microtubule-based, Kuznetsov *et al.*⁴ discovered that membrane vesicles can also move on actin filaments. Intriguingly, some vesicles moved on both microtubules and actin filaments, indicating that they have actin-based (myosin) and microtubule-based (dynein and kinesin) motors on their surfaces. Myosin-based organelle transport may also exist in vertebrates, where mutations in a mouse gene (*dilute*) encoding a myosin-like pro-

tein produce changes in coat colour and in the nervous system that suggest a defect in vesicle transport⁵. Localization studies of this protein confirm that it is associated with vesicular structures in neuronal cells⁶.

A role for microtubule-based vesicle transport in *Saccharomyces cerevisiae* has emerged from studies on MYO2, the yeast homologue of the *dilute* myosin-like protein⁷. Mutations in the MYO2 gene produce defects in polarized secretion, suggesting a role for this actin-based motor in the intracellular transport of vesicles. To identify proteins that regulate or interact with this myosin, genes were isolated that when present in increased copy number partially suppress the MYO2 defects⁸. Surprisingly, one gene (*SMY1*) was recovered that encoded a member of the kinesin superfamily of microtubule motor proteins. Deletion of *SMY1* caused no phenotype, although it did cause synthetic lethality in combination with MYO2 mutations. These results raised the possibility that microtubule systems play some part in polarized secretion, which was previously unrecognized owing to the presence of an active and essential actin-based system.

Why do cells possess both actin- and microtubule-based organelle transport systems? One possibility is that they serve redundant functions in membrane movement, in which case some cells may have become more reliant on one system than the other. Another is that microtubules provide the primary highways for long-range transport between the cell centre and the periphery, whereas the cortical actin network supplies the tracks for the final step of delivering vesicles directly to the plasma membranes⁹.

Although the discovery of potentially redundant actin- and microtubule-based organelle transport within the same cell is perhaps not surprising in retrospect, there was little reason to expect that the components of one transport system would turn up as structural or regulatory components of the other. However, Lees-Miller *et al.*¹, Clark and Meyer², and R. B. Vallee and colleagues (personal communication) have now identified a novel and divergent form of actin associated with elements of the microtubule system. Through different and rather serendipitous means, these groups cloned a gene encoding a protein with ~55% amino-acid identity to conventional actins and ~40% identity to another newly discovered divergent actin (ACT2) from *S. cerevisiae*¹⁰ and *Schizosaccharomyces pombe*¹¹. Unexpectedly, this protein (termed centractin by Clark and Meyer and actin-RPV by Lees-Miller *et al.*) was not detected in actin-rich regions of the cells. Instead, Clark and Meyer discovered by immunocytochemistry and cellular fractionation that this actin is located at centrosomes, whereas Lees-Miller *et al.* found that it is a component of the dynactin complex, which enables cytoplasmic dynein to transport vesicles along microtubules in an *in vitro* assay.

What is the function of actin-related proteins in the centrosome and the dynactin complex? One general possibility is that they have structural roles such as anchoring centrosomes to nuclei or dynein motors to vesicles. Alternatively, they may directly regulate microtubule-based motors (dynein- or kinesin-like) needed for centrosomal or vesicle movements. This possibility is backed up by an earlier finding that actin is a component of *Chlamydomonas* axonemes¹², a structure previously thought to be composed exclusively of microtubules and their associated proteins. Recent findings¹³ suggest that axonemal actin is part of a dynein regulatory complex

1. Lees-Miller, J. P., Helfman, D. M. & Schroer, T. A. *Nature* **359**, 244–246 (1992).
2. Clark, S. W. & Meyer, D. I. *Nature* **359**, 246–250 (1992).
3. Gill, S. R. *et al.* *J. Cell Biol.* **115**, 1639–1650 (1991).
4. Kuznetsov, S. A., Langford, G. M. & Weiss, D. G. *Nature* **356**, 722–725 (1992).
5. Mercer, J. A., Seperack, P. K., Strobel, M. C., Copeland, N. G. & Jenkins, N. A. *Nature* **349**, 709–712 (1991).
6. Espindola, F. S. *et al.* *J. Cell Biol.* (in the press).
7. Johnston, G. C., Prendergast, J. A. & Singer, R. A. *J. Cell Biol.* **113**, 539–551 (1991).
8. Lillie, S. H. & Brown, S. S. *Nature* **356**, 358–361 (1992).
9. Atkinson, S. J., Doberstein, S. K. & Pollard, T. D. *Curr. Biol.* **2**, 326–328 (1992).
10. Schwob, E. & Martin, R. P. *Nature* **355**, 179–182 (1992).
11. Lees-Miller, J. P., Henry, G. & Helfman, D. M. *Proc. natn. Acad. Sci. U.S.A.* **89**, 80–83 (1992).
12. Piperno, G. & Luck, D. J. *J. L. J. biol. Chem.* **254**, 2178–2190 (1979).
13. Piperno, G., Mead, K. & Shestak, W. *J. Cell Biol.* (in the press).
14. Flaherty, K. M., McKay, D. B., Kabsch, W. & Holmes, K. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5041–5045 (1991).

RÉSUMÉ

Plumbing volcanoes

FROM analyses of radioactive decay series in erupted lavas, D. M. Pyle (*Earth planet. Sci. Lett.* **112**, 61–73; 1992) has estimated the residence time of magma in reservoirs beneath Etna, Stromboli and Oldoinyo Lengai, volcanoes in Italy and the East African rift valley. Decay series are a chain of nuclear reactions like a succession of waterfalls pouring at different rates from one pool to the next; unstable parents decay rapidly and nuclide concentrations build up in front of slow reactions causing so-called radioactive disequilibria. Taking volcano plumbing to be a reservoir with balanced input and output, Pyle concludes that magmas reside beneath these volcanoes for as little as 10–100 years.

Count down

WARNINGS that the quality of human semen has declined in recent years are well founded, say E. Carlsen *et al.* in the *British Medical Journal* (**305**, 609–613; 1992). The authors looked at 61 papers that appeared between 1938 and 1990 and contained data from nearly 15,000 men; the geographical spread was wide, though almost half of the studies came from the United States. From their statistical analysis it seems that the mean sperm count dropped from $113 \times 10^6 \text{ ml}^{-1}$ in 1940 to $66 \times 10^6 \text{ ml}^{-1}$ in 1990, the corresponding figures for seminal volume being 3.40 ml and 2.75 ml. Could these apparent declines be due to variation in methods over the years, or bias in the populations sampled? Carlsen *et al.* think not, and point to an increase in certain genitourinary complaints over the same period as a possible cause.

Riding along

MASSIVE black holes are thought to be common in galaxy cores, and at least some galaxies formed through the merging of others. So some galaxies must contain two massive black holes, which will themselves eventually merge. T. Fukushige, T. Ebisuzaki and J. Makino (*Astrophys. J.* **396**, L61–L63; 1992) say that such mergers take a billion years or less, and estimate that every 5 years or so, somewhere in the Universe, two black holes will swallow each other. The coalescence causes a temporary but large twisting of spacetime, whose effects propagate outwards as a burst of gravitational radiation. Precise tracking of interplanetary spacecraft might reveal the passage of the gravity waves, although Voyager 1's trajectory, monitored briefly in 1981, showed nothing out of the ordinary. But with an event expected every few years, Fukushige *et al.* urge more patience.

which may control the force-generating activities of a subset of inner dynein arms and mediate their interactions with radial spokes.

To fulfil their functions, the actin-related proteins might form short filaments or oligomers that could be used as scaffolds to assemble additional regulatory proteins. Another possibility is raised by sequence comparison, which reveals^{1,2} that centractin/actin-RPV is most similar to conventional actin in the ATP-binding and hydrolysis regions, and less so in the regions involved in polymerization. So perhaps these divergent actins do not form polymers like conventional actin, but instead use nucleotide hydrolysis as a switch to control their association with other centrosomal or dynein-associated proteins. In this sense, centractin/actin-RPV may be more similar to HSC70, an even more distant relative that shares a similar

three-dimensional structure with actin¹⁴.

Although information is sparse, the interaction between microfilaments and microtubules promises to be an exciting area of research. Besides the examples described here, interactions of microtubules with the actin cortex determine the position of the cleavage furrow and may be involved in positioning centrosomes and nuclei during the early divisions in development. Such systems should provide fertile grounds for discovering molecules that interconnect and communicate information between the microfilament and microtubule systems. □

Lawrence S. B. Goldstein is in the Department of Cellular and Developmental Biology, Harvard University, Cambridge, Massachusetts 02138, USA. Ron D. Vale is in the Department of Pharmacology, University of California at San Francisco, San Francisco, California 94143, USA.

DIABETES

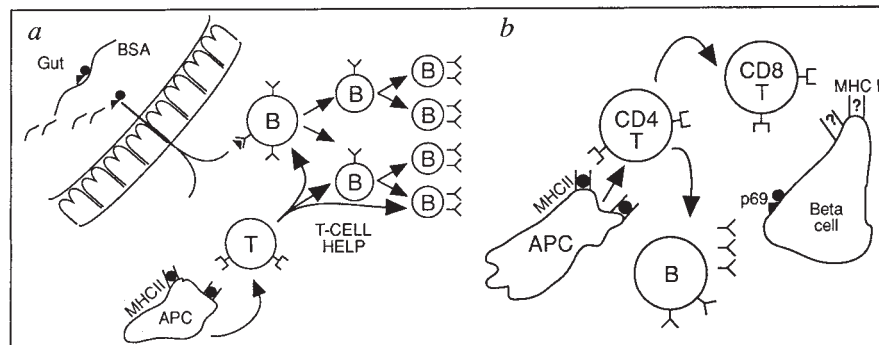
Breast may well be best

Alan G. Baxter and Anne Cooke

ENVIRONMENTAL factors are known to play a role in breaking tolerance to pancreatic β -cell antigens and precipitating the onset of autoimmune (type 1) diabetes mellitus in those individuals genetically prone to develop the disease. But the nature of these antigens, whether they are viral or dietary, remains controversial. Considerable effort has been put into the identification of the β -cell antigens recognized by auto-reactive B or T cells, and several candidate antigens have been identified. Now, Karjalainen and colleagues¹ describe observations supporting their intriguing hypothesis that destruction of the insulin-producing β -cells is caused by an immune response to cows' milk.

Karjalainen *et al.* found that serum

concentrations of antibodies against bovine serum albumin (BSA) were higher in patients with recent onset diabetes than in controls. Patients did not show raised concentrations of antibody to other antigens in cows' milk. Most of the anti-BSA antibody appeared to be directed against a 17-amino-acid section of the molecule (amino acids 152–168) which differs in sequence between humans and cattle. The authors previously showed that immunization with a synthesized peptide from this region of BSA, which they called ABBOS, induced an antibody capable of reacting with a pancreatic β -cell surface protein, p69. Although p69 is not normally expressed, it can be induced to appear on β -cells by the inflammatory mediator, γ -interferon.



a, Bovine serum albumin (BSA) from cows' milk is partially digested within the immature gut and large peptide fragments are absorbed. One such peptide contains an antigenic epitope recognized by T cells (●) and an epitope recognized by B cells (▲). This peptide triggers an immune response and, with T-cell help, B cells produce an antibody that binds BSA. b, Infection subsequently triggers expression of p69 on pancreatic β -cells. By coincidence, p69 contains both the B-cell epitope and the T-cell epitope of the BSA peptide and stimulates a secondary immune response which destroys β -cells. APC, antigen-presenting cell.

CYSTIC FIBROSIS

More from the modellers

James M. Wilson and Francis S. Collins

Further support for the idea that an immune response originally directed against BSA may cross-react with an islet antigen comes from the finding, by the same group², that anti-BSA antibodies found in patients also bind to p69.

Karjalainen and colleagues offer a hypothesis which is elegant and simple. The ABBOS peptide is immunogenic in hosts with a diabetogenic haplotype, and is absorbed from the immature gut. The resulting ABBOS-specific immune response could mediate destruction of β -cells through the "shared epitope", p69. Unrelated infectious events that generate the release of γ -interferon would induce expression of p69, transiently exposing β -cells to immune-mediated destruction (see figure). The long preclinical course preceding the onset of diabetes could be explained by the temporary nature of such episodes of p69 expression on β -cells.

There are, however, some problems inherent in this study as it stands. The authors boldly state that ABBOS and p69 share an antigenic epitope based on cross-reactive antibody binding. As destruction of β -cells is believed to be primarily T-cell mediated, T-cell reactivity is of greater significance than antibody specificity and, in any case, would be necessary for a secondary anti-p69 antibody response. Therefore Karjalainen *et al.* must imply that p69 and BSA contain a homologous linear peptide sequence in order to trigger the T-cell response. A further difficulty is that binding of antibody to a molecule does not necessarily indicate biological relevance, as it says nothing of binding affinity or ability to mediate effector functions³. These points are not just academic. Diabetic patients have long been known to have a variety of auto-antibodies, including those against insulin⁴ and a 64K islet-cell protein⁵, and although they are of predictive value they have failed to tell us much about the pathogenesis of β -cell destruction.

So, intriguing though it is, the hypothesis requires further testing. It remains to be seen whether ABBOS and p69 do have a similar amino-acid sequence, whether the T cells of diabetics are responsive to p69, and whether reactivity to p69 is pathogenic, or results from spreading of autoimmunity to secondary antigens following tissue destruction. □

Alan G. Baxter and Anne Cooke are in the Department of Pathology, Division of Immunology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK.

THE mouse model for cystic fibrosis published last month^{1,2} (and discussed by us in News and Views³) was a remarkable scientific achievement and has had an immediate influence on investigations of this deadly disease. That model was developed by a group at the University of North Carolina (UNC) using homologous recombination in embryonic stem cells. On page 211 of this issue⁴, David Porteous from Edinburgh and his collaborators (Dorin *et al.*) now



A litter of mice resulting from the mating of two CF insertional mutant heterozygotes; the litter consists of five wild type (+/+), three heterozygotes (*cf*/+) and one CF insertional mutant homozygote (*cf*/*cf*). The *cf*/*cf* mutant is the black mouse. Coat colour variation is independent of the CF gene, but the animal's genotype was identifiable by electrophysiology and histological abnormalities. (Picture courtesy of J. R. Dorin *et al.*)

announce the generation of another murine model of cystic fibrosis with embryonic stem cell technology. This new model expresses a surprisingly different phenotype to that found in the UNC animals.

In our previous News and Views article we described the salient features of the cystic fibrosis (CF) model from the UNC group. Animals homozygous for the disrupted CF gene, subsequently referred to as *cf*/*cf*, develop an extremely reproducible phenotype characterized by death within 30 days of birth due to bowel obstruction, a condition that closely resembles the syndrome of meconium ileus in humans. Pathology of the bowel, nasal mucosa and submaxillary glands was characteristic of the disease, but there was a notable absence of symptoms in the lungs, pancreas, hepatobiliary system and reproductive tract.

The *cf*/*cf* mouse described in this issue differs in several ways. Evaluation of the new model was restricted to eight *cf*/*cf*

mice for a period up to 30 days postpartum. Electrophysiological abnormalities were identified in epithelia of each animal consistent with defective transport of chloride ions. Unlike the UNC animal model, however, these *cf*/*cf* mice developed a variable spectrum of pathological features and remained free of overt clinical disease. Most animals had characteristic pathology in gut mucosa, but two of the eight *cf*/*cf* animals were indistinguishable from non-CF animals in terms of gross and histological pathology. In addition, the authors describe focal pathology in lung, vas deferens and salivary glands of single animals; the relevance of these isolated observations to the expression of CF in this model will require longitudinal blinded studies with larger numbers of animals.

Several mechanisms could explain the variation in phenotype observed between the two models. Both of them were developed from a line of embryonic stem cells in which the gene encoding the cystic fibrosis transmembrane conductance regulator (CFTR) had been disrupted at exon 10 by homologous recombination. The UNC group used a replacement vector to in-

troduce an in-frame stop codon in the coding sequence of exon 10, whereas the Edinburgh group used an insertional vector to introduce a duplication of genomic sequences spanning intron 9 and exon 10. An advantage of the insertional strategy is that it can be used in subsequent steps to introduce specific mutations into the CF gene such as the common $\Delta F508$ mutation (in fact Dorin *et al.* state that such a cell line has been generated). The disadvantage is that the transcript derived from this kind of gene disruption may splice out the duplication at a low level thereby reconstituting small amounts of normal CF gene function. The milder phenotype of the *cf*/*cf* animal described by Dorin *et al.*, therefore, may be explained by low levels of residual CFTR function. Other causes of phenotypic heterogeneity between and within *cf* genotypes include environmental influences and the genetic background into which the *cf* mutation has been bred (into three different inbred strains in the study by Snouwaert *et al.*¹

1. Karjalainen, J. *et al.* *New Engl. J. Med.* **327**, 302–307 (1992).

2. Martin, J. M. *et al.* *Ann. Med.* **23**, 447–452 (1991).

3. Bankhurst, A. D. & Williams, R. C. J. *clin. Invest.* **56**, 1378–1385 (1975).

4. Palmer, J. P. *et al.* *Science* **222**, 1337–1339 (1983).

5. Baekkeskov, S. *et al.* *Nature* **298**, 167–169 (1982).

or into an outbred strain in the study by Dorin *et al.*⁴).

This new animal model of CF has several features that will greatly enhance its usefulness. The apparent absence of lethal bowel obstruction may allow survival until pathology in other organ systems such as lung, pancreas or liver can be further elicited. In addition, the marked variation of phenotype in animals with the same *cf* genotype provides an excellent opportunity to identify environmental and additional genetic factors that contribute to expression of disease. This heterogeneity in phenotypic expression, however, will need to be taken into account in studies of potential therapies that use clinical or pathological endpoints.

The apparent differences in the first two mouse models of CF underscore the complexities of modelling human dis-

eases in animals. They are a timely reminder of the value of having important problems tackled by more than one research group, even if the differences in strategy seem minor. The challenge remains to use the full power of mouse genetics to understand the basis of the differences between these models in the context of the pathophysiology of CF in humans. □

James M. Wilson and Francis S. Collins are in the Howard Hughes Medical Institute, the Division of Molecular Medicine, and the Cystic Fibrosis RDP Center at the University of Michigan, 4570 MSRB II, Box 0650, Ann Arbor, Michigan 48109, USA.

1. Snouwaert, J. N. *et al. Science* **257**, 1083–1088 (1992).
2. Clarke, L. L. *et al. Science* **257**, 1125–1128 (1992).
3. Collins, F. S. & Wilson, J. M. *Nature* **358**, 708–709 (1992).
4. Dorin, J. R. *et al. Nature* **359**, 211–215 (1992).

PALAEOCLIMATOLOGY

Avoiding a permanent ice age

William R. Kuhn

ALTHOUGH the Earth's climate is clement, on our neighbouring planets things are different. The greenhouse effect has run away on Venus, giving the planet a scorching atmosphere. And Mars is a frozen world. How lucky are we? Calculations by Caldeira and Kasting on page 226 of this issue¹ suggest that climate cooling when the Earth was young would have triggered the growth of carbon dioxide ice clouds high in the atmosphere, locking the Earth into a permanent ice age. The present warmth of our climate might be taken to indicate that previous notions of an early ice age are wrong.

One of the vexing problems of palaeoclimate is known as the faint young Sun paradox². Models of solar evolution and observations of stars like our Sun in various stages of evolution indicate that solar luminosity was significantly less — by 25–30 per cent — than it is today — so much so that the Earth would have been a frozen world if all else remained the same. Yet there is geological evidence that there was liquid water on the planet up to 3.8 billion years ago³. Speculation has centred on increased concentrations of certain gases in the atmosphere that create a greenhouse effect, most notably carbon dioxide, and arguments have been put forth that the resulting warming could counteract the faint young Sun and save the planet from total glaciation.

Although the earlier studies of palaeoclimate recognized that carbon dioxide, in large enough amounts, could have counteracted the lower solar lumi-

nosity, they did not include the possibility that carbon dioxide ice crystals could have formed. Caldeira and Kasting speculate that air temperatures could have fallen enough for carbon dioxide ice clouds to have appeared in the young atmosphere. These clouds might have looked much as cirrus water ice clouds do today, which frequently appear as white, wispy or filamentary clouds above about 6 km. But there is a substantial difference: water ice clouds generally heat the Earth because of their substantial greenhouse effect⁴, whereas carbon dioxide ice clouds would have more of a tendency to cool the surface because they are good scatterers of radiation, including solar radiation.

The possible existence of carbon dioxide clouds in the early atmosphere has implications for what are known as run-away effects. Could some transient change in the Earth or its atmosphere have made the early climate unstable, causing polar ice sheets to grow and cover the Earth? If so, then obviously the process was reversible.

Simple models have been developed that indicate the number and stability of climate states for different solar luminosities, and limit the ranges of those parameters that affect climate. For example, studies have shown that decreases of solar luminosity of perhaps only a few per cent could cause polar ice to spread over the whole planet^{5,6}. In the light of Caldeira and Kasting's work, it seems unlikely that this happened even with a 25–30 per cent difference in solar luminosity. Carbon dioxide clouds may

have formed that would have prevented the Earth from ever returning from its frozen state. (Even the build-up of gaseous carbon dioxide — cut off by ice from the Earth's rocky surface, a carbon dioxide sink — would not have negated this.) If Caldeira and Kasting are correct, then the Earth may have been warm from its birth and remained warm with an abundance of greenhouse gases such as carbon dioxide and water vapour.

Our recent awareness that we have the potential to alter global climate has accelerated the interest in and importance of climate modelling. Much of what we learn from those studies is applicable to our efforts to understand what Earth was like billions of years ago. However, we are finding that the Earth system is vastly complex, and it is not obvious whether the relative importance of a process in the geological past was the same as it is today.

For example, the biosphere plays a strong role in regulating climate today, but could hardly have done so a few billions of years ago. And what fraction of the Earth was covered by land? Rapid continental growth occurred 2.7–3 billion years ago⁷, so that before then, there was probably little exposed land. How would this have affected the change in atmospheric carbon dioxide through rock weathering?

Solar radiation is clearly the most important determinant of the Earth's temperature. Last year, Graedel *et al.* called into question the magnitude of the solar flux early in Earth's history⁸. It is possible that the young Sun was not as faint as we thought. New solar models, albeit speculative, include an early solar mass loss that leads to luminosities larger than the 75–80 per cent generally assumed.

Another important uncertainty in modelling both palaeoclimate and present climate is the treatment of clouds. If it weren't for clouds in the present atmosphere, the temperature would be some 20 °C higher. Most clouds in our present-day atmosphere are more effective at reflecting solar radiation away from the surface than they are at trapping heat by the greenhouse effect. Cirrus clouds are the exception. It is not only cloud amount, thickness and droplet size that are important, but also

1. Caldeira, K. & Kasting, J. F. *Nature* **359**, 226–228 (1992).
2. Newman, M. J. & Rood, R. T. *Science* **198**, 1035–1037 (1977).
3. Sagan, C. & Mullen, G. *Science* **177**, 52–56 (1972).
4. Liou, K.-N. & Gebhart, K. L. *J. met. Soc. Japan* **60**, 570–582 (1982).
5. Sellers, W. D. *J. appl. Met.* **8**, 392–400 (1969).
6. Sellers, W. D. *Palaeogeogr. Palaeoclimat. Palaeoecol.* **82**, 217–224 (1990).
7. Condie, K. C. *Palaeogeogr. Palaeoclimat. Palaeoecol.* **75**, 57–81 (1989).
8. Graedel, T. E., Sackmann, I. -J. & Boothroyd, A. I. *Geophys. Res. Lett.* **18**, 1881–1884 (1991).

height, which determines whether or not the temperature is low enough for ice clouds to form. The colder temperatures that would cause carbon dioxide ice crystals to form in the primitive Earth atmosphere might also cause water ice clouds to increase. If so, the cooling from carbon dioxide ice would have been counteracted. This is but one of the many unanswered questions that need to be explored.

In any case, as we improve our know-

ledge of the present Earth system we will gain further insight into the Earth's early climate. And while there is little for certain that we can say about that climate today, we should be able to narrow the bounds within which our climate evolved. □

William R. Kuhn is in the Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA.

GEOPHYSICS

Earth's core not so hot

Al Duba

THE Earth's core is primarily iron, with a solid centre and a liquid outer region which is in contact with a solid mantle. The density and moment of inertia show that there is also a lighter element, sulphur or oxygen, alloyed with the iron. Calculating the temperature in this deep zone has been a time-honoured pastime in geophysics, because accurate knowledge of it is important for constraining both the Earth's radioactive heat budget and the physical conditions in which its magnetic dipole is generated. In a recent paper¹ fuel is added to the fires of controversy raging within the 'ironworker' community concerning temperatures in the Earth's core.

Using a laser-heated diamond-anvil cell (DAC), Boehler¹ and colleagues² from Mainz, Germany, measure melting points of Fe, FeO and FeS, the three most likely constituents of the Earth's core, to pressures of 114, 47.0 and 43.5 gigapascals (GPa), respectively. From these data a minimum temperature for the core-mantle boundary (CMB) is derived, ranging from 2,600 to 3,400 K. This result is in sharp contrast to the 3,600–5,300 K minimum temperature range proposed for the CMB based on Fe and FeO melting-point data reported earlier by Jeanloz and colleagues^{3,4} from Berkeley, United States (see table). These differences become even larger at the inner-core/outer-core boundary where Earth's molten outer core contacts the solid inner core.

It is at this boundary that there is the possibility of setting a theoretical upper limit on the temperature if one makes

the first-order assumption that Earth's core is pure iron and allows about 1,000 K for possible eutectic melting caused by the presence of the lighter element. Unfortunately, this theoretical calculation is subject to large uncertainties because it relies on the density dependence of thermodynamic properties which are difficult to obtain experimentally⁵. The inner-core/outer-core boundary temperature from theoretical consideration is intermediate to those extrapolated from the two experimental groups with a shading of a few hundred degrees in favour of the Berkeley results.

Figure 1 presents the new melting data of Boehler¹ which are in good agreement with those reported for FeO as measured in a multi-anvil press⁶ and in poor agreement with those obtained by the Berkeley group³ in a laser-heated DAC. And Figure 2 compares the melting of Fe determined by the Mainz² and Berkeley⁴ groups in the laser-heated DAC. The disparity between these two groups is astounding and reflects the difficulties of recreating core conditions in the laboratory. From the Figures and from the experimental procedures detailed by the authors, it seems unlikely that pressure measurement is responsible for the disagreement, as both groups use ruby fluorescence, a technique that has been tested and calibrated *ad nauseam*.

The rub probably lies in the melting-point measurement itself, which has two components — the observation of melting, and temperature determination concurrent with melting. The Figures reveal

that the data scatter of Boehler and his colleagues is considerably less than that of the Berkeley group. Although it is not in itself an indication of the relative merits of two conflicting measurements, data scatter can emphasize the difficulties inherent in a particular measurement technique.

The Berkeley group uses four criteria to determine melting⁷. During the experiment, visual observations of textural changes on the surface of the sample following laser heating over the entire pressure range and of fluid motion within the sample during laser-heating at pressures up to 50 GPa were used. These observations are followed up, post-experiment, by optical examination of all samples and electron-microscopic examination of about half the samples. The temperature of the melt zone is calcu-

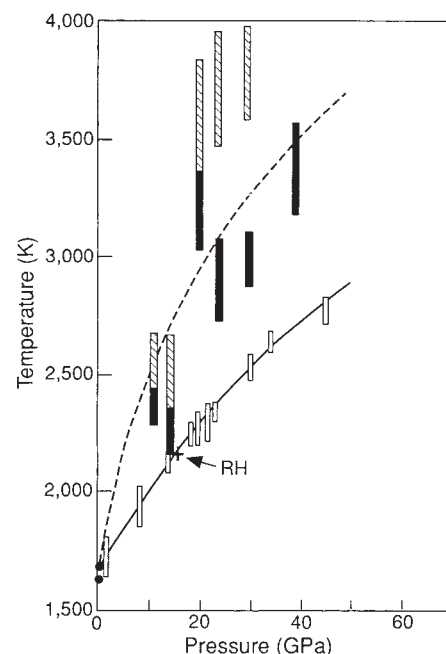


FIG. 1 Melting of FeO in the diamond-anvil cell. Open bars indicate variation of melting temperature observed¹ in at least four different measurements of onset of melting in Fe_{0.96}O. The cross that is marked RH is for Fe_{0.97}O from Ringwood and Hibberson⁶ and the data from Knittle and Jeanloz³ for Fe_{0.94}O are represented by the hatched (melt) and filled (solid) bars.

lated from average temperatures measured spectroradiometrically⁸. The melt zone is typically 2–4 µm in diameter, the laser beam doing the heating is 20–30 µm in diameter, and the field of view of the spectrometer is 90 µm in diameter. Average temperatures of the laser-heated sample area are determined from the spectrometer readings using Wien's approximation to Planck's law.

The temperature gradient across the laser-heated spot is then determined by scanning a 25-µm-wide sampling slit across a 250–300-µm-diameter image of

CORE TEMPERATURES DERIVED WITH SOLID-SOLUTION AND EUTECTIC MODELS FROM DAC DATA

Source	Core-mantle boundary		Outer-core/inner-core boundary	
	Solid-solution	Eutectic	Solid-solution	Eutectic
Mainz group ^{1,2}	3,200–3,400 K	2,600–2,800 K	4,200 K	3,500–3,800 K
Berkeley group ^{3,4}	4,800±500 K	3,800±200 K	7,600±500 K	6,600±500 K
Grover ⁵	—	—	—	5,500 K

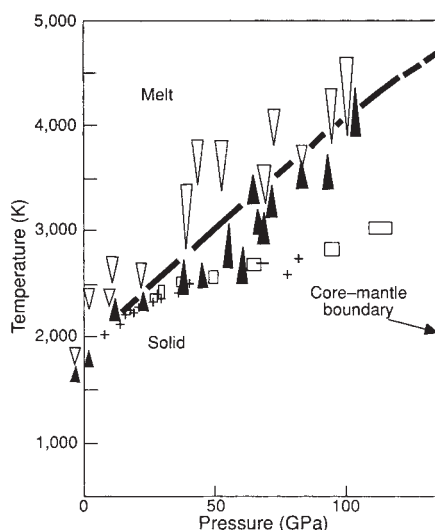


FIG. 2 Melting of Fe in the diamond-anvil cell. The crosses represent laser-power/temperature-discontinuity indications of melt and rectangles are reflected-light indications of melt². Size of symbols is the uncertainty in temperature and pressure determination. Triangles are from ref. 4. The open downward-pointing triangles represent the lowest temperatures measured for melting. Solid upward-pointing ones indicate highest temperature observed in solid. Length of symbol is an indication of the temperature uncertainty.

the heated spot. The profiles of intensities (at four wavelengths between 600 and 750 nm) as a function of distance across the sample are inverted to yield intensity of light as a function of distance from the centre of the hot spot. Then Wien's approximation is used to determine the temperature distribution across the sample (and hence, the temperature at the central 2–4 μm where melting occurred).

Despite the approximations, inversions and least-squares fitting, the authors claim an uncertainty of 20–50 K for the melting-point temperature. Because the calculation of the melting temperature depends on processing optical data through the Wien approximation twice in reciprocal temperature space where temperature errors are nonlinear as temperature increases, this claimed error seems very optimistic, especially at the higher temperatures. Uncertainties of hundreds of degrees seem the more likely outcome.

The Mainz group has a larger molten area in their DAC (between 5 and 10 μm in diameter) and they use an optical arrangement to permit collecting emission spectra from 2- μm -diameter areas within the melt zone. Thus, the data are processed through the Planck radiation function only once and, in addition, there is no inversion to obtain an average temperature of the laser-heated zone within a 90- μm aperture as is required for the Berkeley group. By measuring

several 2- μm -diameter zones within the 5–10 μm -diameter melt zone, the Mainz group claims a temperature gradient below 15 K μm^{-1} within that melt zone.

Melting is detected by two different methods *in situ*: when melting occurs, a sharp discontinuity is seen in the laser-power/optical-temperature relationship, accompanied by formation of either a hole or a pit, with beaded rims in the iron foil, and the intensity of 488-nm laser light reflected from the hot spot drops sharply.

Estimates of core temperatures from these two groups from various publications^{1–4} are assembled in the table. Because the Mainz group calculates its temperature from emission spectra collected within the melt zone, whereas the Berkeley group inverts data collected over an area nine or more times larger than the laser beam and 500 or more times larger than the melt zone, I conclude that the melting temperature of Fe and FeO is nearer to that measured by the Mainz group.

However, the question of the temperature of the core is not so easily resolved. Boehler¹ uses the Kraut-Kennedy melting relationship⁹ to extrapolate his data to pressures at the CMB and inner-core/outer-core boundary. Kraut and Kennedy pointed out that there was no theoretical basis for their melting relationship and many in the community involved in high-pressure equation-of-state calculations feel it extremely unreliable for large volume compression in non-alkali metals (M. Ross, personal communication). The very low CMB temperature of the Mainz group implies that the core may not contribute to the Earth's heat-flow budget. Perhaps the dynamo responsible for the magnetic field derives its energy from radioactive heating supplied by the mantle. Perhaps the heat generated, as other ironworkers weigh in over the coming months and as the Mainz and Berkeley groups discuss their hot and hotter cores, will be accompanied by a little more light. □

Al Duba is at the Lawrence Livermore National Laboratory, University of California, Livermore, California 94551, USA.

- Boehler, R. *Earth planet. Sci. Lett.* **111**, 217–227 (1992).
- Boehler, R., von Bagen, N. & Chopelas, A. *J. geophys. Res.* **95**, 21731–21737 (1990).
- Knittle, E. & Jeanloz, R. *J. geophys. Res.* **96**, 16169–16180 (1990).
- Williams, Q., Jeanloz, R., Bass, J., Svendsen, B. & Ahrens, T. *J. Science* **236**, 181–182 (1987).
- Grover, R. *J. geophys. Res.* **95**, 21743–21748 (1990).
- Ringwood, A. E. & Hibberson, W. *Phys. Chem. Miner.* **17**, 313–319 (1990).
- Williams, Q., Knittle, E. & Jeanloz, R. *J. geophys. Res.* **96**, 2171–2184 (1991).
- Heinz, D. L. & Jeanloz, R. in *High Pressure Research in Mineral Physics*, Geophys. Monogr. Ser. 39 (eds Manghnani, M. H. & Syono, Y.) 113–127 (Am. Geophys. Un., Washington, DC, 1987).
- Kraut, E. A. & Kennedy, G. C. *Phys. Rev. Lett.* **16**, 608–609 (1966).

DAEDALUS

Sensitive skin

THE best dancers, acrobats and experimentalists have a marvellous skill and dexterity. The awkward and clumsy among us, the trippers over and the smashers of apparatus, can only envy them: for it is not at all obvious what they have that we do not. Daedalus reckons that physical poise and grace come from a just and accurate proprioceptive sense: that subtle sense which gives information about the disposition of the body. The physically clumsy, he claims, do not lack this sense: they simply have not learnt to be aware of it. Instead of sensing their disastrous performance, they merely read their elegant physical intention. In the same way, a speaker is often unaware of a marked verbal accent: instead he 'hears' internally the clear dictation of his intended speech.

So Daedalus is planning a set of proprioceptive aids to do for balance and poise what spectacles do for sight. DREADCO technicians are devising patches to be stuck on the skin, made of a piezoelectric polymer such as polyvinylidene difluoride. They generate a voltage pattern whenever the skin stretches in a particular direction. Near knees and elbows, for example, they act as bend encoders; applied circumferentially around arms and legs they are expanded by shortening muscle and act as force encoders. They are rate sensitive, too. A rapid change generates a stronger current, thus giving differential information.

Electric shock is normally unpleasant and uninformative. But the closely spaced electrode surfaces of Daedalus's skin patches do not launch currents deep into the tissues. Their effect is confined to the skin surface, where it stimulates stretch and hair displacement sensors to give a form of fake contact sense. The DREADCO team should soon discover which electrical patterns are most comprehensible to the wearer and relate most instinctively to his bodily position and rates of movement. He will be puzzled by the sensations at first, but will soon learn to 'ride' his own body, just as one learns to ride a bicycle. The experience should be as intimately delightful as that of a young child discovering the skills of walking and balancing.

Thus the stumblers and butterfingers of society will acquire the grace of the born gymnast and the dexterity of the born experimenter. Their body language will gain clarity and eloquence, and they will at last feel at home in the physical world. The skin patches which bring this liberation will slowly lose adhesion and need to be replaced, but Daedalus is already working on a system of permanent, invisible piezoelectric tattoos.

David Jones

Excrement analysis by PCR

SIR — Samples from endangered animals are hard to obtain for genetic analysis. To study a threatened bear population in the Pyrenees, Taberlet and Bouvet¹ used hair collected from wire netting attached to trees on which the bears scratch themselves. We have approached the problem by the use of bear droppings, which can be collected without disturbing the animals and can be used to amplify DNA sequences not only from the animals themselves but also from the foods they have ingested.

We are studying a population of European brown bears in the Brenta region of Northern Italy. Like the Pyrenean bears, this population is in decline and today may number less than 10 individuals. The introduction of animals from the large populations in the Balkans is contemplated. To evaluate the extent of genetic variation in the Brenta bears as well as their genetic relationship to Romanian bears, we looked at the sequence variation in mitochondrial DNA from the Italian bears. Nucleic acids were extracted from excremental remains, presumed from their morphology to be of bear origin. All extracts contained large amounts of bacterial DNA.

Two oligonucleotide primers were designed to span an 88-base-pair-long segment of the mitochondrial control region

which has been sequenced in the American brown bear². Using these primers in the polymerase chain reaction (PCR), we were able to amplify mitochondrial DNA from three droppings (Fig. 1a). Sequences from the Brenta droppings as well as from a liver sample of a Romanian bear are shown in Fig. 1b. The European bears differ at seven positions from the American brown bear. The three Brenta DNA samples are identical to each other and differ at three positions from the Romanian bear sample. Further work will clarify whether this indicates that the Brenta population is genetically depauperate compared with other bear populations. Statistical analysis will have to take into account that the samples are being drawn with replacement, because individual animals may be sampled several times.

We also used primers specific for the chloroplast *rbcL* gene to investigate whether nucleic acids of plants ingested by the bears can pass through the digestive tract. As can be seen in Fig. 1a, all three bear droppings produced bands. The direct sequencing of the amplification products showed that they contain one *rbcL* sequence. When this sequence was compared with 414 *rbcL* sequences (M. W. Chase, personal communication), it turned out to be identical to *Photinia*, a genus in the *Rosaceae*. A

representative of this genus, *P. villosa*, is found in the Brenta region. Consequently, the fruits of this plant seem to be a dominant component of the bears' diet during late summer when the droppings were collected. Thus the feeding behaviour of animals can be investigated after amplification by PCR of DNA from their excrement. The collection of droppings may also be a useful way to screen large areas for the presence of rare animals.

Matthias Höss

Michael Kohn

Svante Pääbo

Institute of Zoology,

Luisenstrasse 14,

University of Munich,

W-8000 Munich 2, Germany

Felix Knauer

Wolfgang Schröder

Munich Wildlife Society,

Linderhof 2,

W-8107 Ettal, Germany

1. Taberlet, P. & Bouvet, J. *Nature* **358**, 197 (1992).

2. Shields, G. F. & Kocher, T. D. *Evolution* **45**, 218–221 (1991).

3. Pääbo, S., Gifford, J. A. & Willson, A. C. *Nucleic Acids Res.* **16**, 9775–9787 (1988).

4. Boom, R. et al. *J. clin. Microbiol.* **28**, 495–503 (1990).

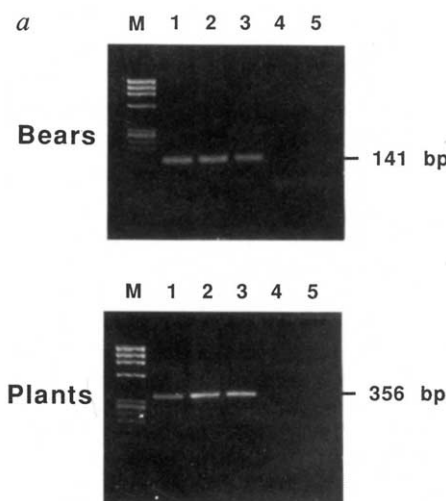
5. Bachmann, B., Lücke, W. & Hunsmann, G. *Nucleic Acids Res.* **18**, 1309 (1990).

Evidence for some hot dark matter?

SIR — The recent detection of the large scale microwave background anisotropy by the cosmic background explorer (COBE) satellite is in remarkable agreement with the predictions based on an inflationary scheme in which the dark matter is a mixture of cold and hot components. We briefly summarize how the COBE measurements, inflation, and the observed large-scale structure all seem to point to a flat Universe in which a fraction (~20–30 per cent) of the critical energy density resides in massive (~3–9 eV) relic neutrinos. A crucial test of this scheme comes from the anisotropies expected on smaller (~1°–2°) angular scales, which are estimated to lie within reach of current and planned experiments.

A generic prediction of inflation is that the Universe contains a critical density of matter. Primordial nucleosynthesis arguments imply that <90% of this matter is nonbaryonic, but inflation does not specify what the dark matter should be. Typically, investigators have assumed the dark matter takes one of two possible forms: cold dark matter (CDM; massive cold particles) or hot dark matter (HDM; relic light neutrinos). Although pure HDM has several well-discussed problems, models with pure CDM provide a basis for structure

a, PCR amplifications of bear mitochondrial DNA from three droppings. Lane M, size marker (ΦX174 DNA digested with the restriction enzyme *Hae*III); lanes 1–3, three droppings from the Brenta population; lanes 4 and 5, extraction and PCR controls, respectively. Upper panel, amplification of a 141-base-pair (bp) fragment (including primers) from the mitochondrial control region. Lower panel, amplification of a 356-bp fragment (including primers) of the chloroplast *rbcL* gene. b, Nucleotide sequences of the 88-bp mitochondrial fragment from three individuals of the Brenta population (Brenta 1, 2 and 3). Sequences are compared with the American brown bear² and a brown bear from Romania. Droppings (0.5 g) were extracted as described in ref. 3, except that the phenol/chloroform and concentration steps were replaced by silica purification⁴. For PCR, we used 25-μl reaction volumes



containing BSA (2 μg mg⁻¹) to overcome inhibition and 10 μl (10%) extract. Each of 40 PCR cycles consisted of denaturation at 92 °C for 40 s, annealing at 55 °C for 60 s and extension at 72 °C for 60 s. Primers for the mitochondrial fragment: 5'-CGTGCATTATGGCGTGCCCA-TGCAT-3' and 5'-TGGTGATCAAGCTCCCGGACTAAGTG-3', and for the *rbcL* fragment: 5'-ATG-TCAACCAAAACAGAACTAAAGCAAGT-3' and 5'-CCAAGACAATGATTGAACAATACTTC-3'. Amplified DNA was electrophoresed on 2.8% low-melting agarose, the bands were cut out and melted in 100 μl water and then reamplified in 50 μl using the same conditions, except for a 60 °C annealing temperature. The products of this second amplification were purified using GeneClean (Dianova) and directly sequenced⁵.

containing BSA (2 μg mg⁻¹) to overcome inhibition and 10 μl (10%) extract. Each of 40 PCR cycles consisted of denaturation at 92 °C for 40 s, annealing at 55 °C for 60 s and extension at 72 °C for 60 s. Primers for the mitochondrial fragment: 5'-CGTGCATTATGGCGTGCCCA-TGCAT-3' and 5'-TGGTGATCAAGCTCCCGGACTAAGTG-3', and for the *rbcL* fragment: 5'-ATG-TCAACCAAAACAGAACTAAAGCAAGT-3' and 5'-CCAAGACAATGATTGAACAATACTTC-3'. Amplified DNA was electrophoresed on 2.8% low-melting agarose, the bands were cut out and melted in 100 μl water and then reamplified in 50 μl using the same conditions, except for a 60 °C annealing temperature. The products of this second amplification were purified using GeneClean (Dianova) and directly sequenced⁵.

formation which fits galactic observations quite well. The amplitude of the CDM density fluctuations had to be biased so it would yield fluctuations in mass which were a factor $1/b$ times smaller than fluctuations in galactic number density (evaluated on a scale of $8H^{-1}$ Mpc, where H is the Hubble constant in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$, and b is bias parameter). N-body simulations¹ of a universe filled with cold particles showed that the best fit between CDM overdensities and galaxies was obtained when $1.7 < b < 3.0$. But with this level of bias, CDM underpredicts the amount of large-scale structure characterized by such quantities as cluster correlations, bulk flows of galaxies, and galactic concentrations observed in redshift surveys. Furthermore, it predicts a quadrupole signal of $4.7 \times 10^{-6}/b$, which is significantly below the measured COBE value² of $6.1 \pm 1.5 \times 10^{-6}$.

By 1983, grand unified theories of particle physics had been constructed^{3,4} in which both types of dark matter (cold plus hot dark matter, or CPHDM) arose naturally with comparable cosmological densities. The formation of large-scale structure with CPHDM was sketched out⁵, and there were more precise calculations of the power spectra. It was emphasized that CPHDM gave more large-scale power and less small-scale power than a similarly normalized CDM model.

The CPHDM models were also more consistent^{6,7} with bulk flow velocities. A consequence of their enhanced large-scale power was that the prediction^{6,7} for the quadrupole moment was somewhat bigger than for the CDM model. The HDM fraction and the bias parameter were determined from comparisons of CPHDM models with observations⁸. A model with $1/4$ HDM and $3/4$ CDM and a realistic bias parameter of about $b=1.1-1.4$ gives a good fit to cluster number densities, cluster correlations, bulk flows of galaxies, and the Infra-red Astronomy Satellite survey power spectrum. We recently showed⁹ that CPHDM could account for the estimated co-moving density of high redshift quasars out to a redshift of 6. For this

model, the quadrupole is predicted to be $7.8 \times 10^{-6}/b$, which for $b=1.3$ is in excellent agreement with the COBE value.

In short, CPHDM models with roughly $1/4$ HDM and $3/4$ CDM score well on every test applied to them. Indeed, they do very well at explaining galaxy formation¹⁰, the traditional stronghold of CDM models. Perhaps the most crucial new test for the CPHDM model will come from measurements of the anisotropies expected on smaller angular scales. The predicted values of the anisotropy (with $H=0.5$ and 5% baryons) on 1° and 2.1° are $3.4 \times 10^{-5}/b$ and $2.3 \times 10^{-5}/b$, respectively. For $b=1.3$, the 1° prediction is less than a factor of two below the current upper limit¹¹.

Robert K. Schaefer

Qaisar Shafi

*Bartol Research Institute,
University of Delaware,
Newark, Delaware 19716, USA*

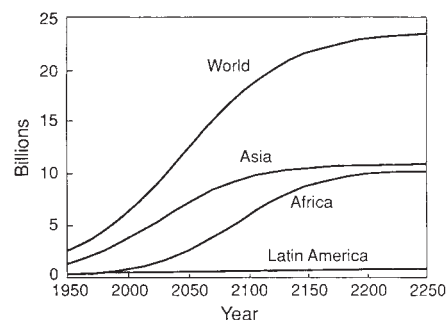
World population

SIR — The United Nations makes predictions of the human population of the Earth and continental regions¹ by the component method². This requires a large number of parameters, such as birth and death rates within various age groups. An alternative approach is to use a model based on a differential equation involving only a few parameters. Three well-known such models are the malthusian, Verhulst's logistic and the gompertzian. The malthusian has been used to estimate previous global populations³, but the logistic and gompertzian have the desirable property of limiting the final population size. We have examined human population data in relation to these three models.

The logistic model was used early this century to examine population trends in the United States. Interest in this model was revived when the populations of England, Scotland and the United States were well-fitted by logistic curves⁴, but a jump in the parameters occurred just after the Second World War. The same phenomenon was found for Australia and New Zealand⁵ at the same time. Thus there are apparently logistic-type regimes which persist until some major event occurs.

The parameters of the best-fitting curves for the various models were computed with a nonlinear least-squares algorithm⁶. The logistic provided an excellent fit to the world and continental population data available at 5-year intervals from 1950 to 1985 (ref. 1). The predicted world population for 1992 is 5.48 billion, in agreement with the UN figure of April 1992.

If the present logistic regime persists,



World and major continental population trajectories to the year 2250 obtained from the best-fitting logistics.

the world population will double in 47 years and the eventual world population will be 23.8 billion. This will be practically attained by the year 2200. The estimates of the final continental populations are as follows: Africa, 10.57 billion; Asia, 11.08 billion; Commonwealth of Independent States, 320 million; Europe, 556 million; Latin America, 899 million; North America, 336 million; and Oceania, 40 million. The figure shows the predicted population trajectories for the world and major continental regions. In all cases the predictions are much higher than those of the United Nations.

The malthusian and gompertzian models also gave good fits to the data to 1985. But the malthusian gives rise to an infinite population and the gompertzian predicts that saturation will occur in a few millennia at a final population of 1,062 billion, which seems unrealistic. (Note that the gompertzian is used mainly in biology to describe the development of masses of cells, as in certain organisms, their organs or tumours⁷.)

The logistic has been derived as a model for populations which may disperse throughout a finite habitat⁸. It will be of interest to see when saturation effects are apparent in the world population, as they are already in the Commonwealth of Independent States, Europe and North America.

Henry C. Tuckwell

*Statistics Research Section,
Australian National University,
GPO Box 4, Canberra 2601, Australia*

James A. Koziol

*Department of Molecular and
Experimental Medicine,
The Scripps Research Institute,
10666 N. Torrey Pines Rd, La Jolla,
California 92037, USA*

1. Frenk, C. S., White, S. D. M., Efstathiou, G. & Davis, M. *Astrophys. J.* **351**, 10–21 (1990).
2. Smoot, G. F. et al. *Astrophys. J.* (in the press).
3. Holman, R., Lazarides, G. & Shafi, Q. *Phys. Rev. D* **27**, 995 (1983).
4. Mohapatra, R. N. & Senjanovic, G. *Z. Phys. C* **17**, 53 (1983).
5. Shafi, Q. & Stecker, F. W. *Phys. Rev. Lett.* **53**, 1292–1295 (1984).
6. Schaefer, R. K., Shafi, Q. & Stecker, F. W. *Astrophys. J.* **347**, 575–589 (1989).
7. Holtzman, J. *Astrophys. J. Suppl.* **71**, 1–24 (1989).
8. van Dalen, T. & Schaefer, R. K. *Astrophys. J.* (in the press).
9. Schaefer, R. K. & Shafi, Q. TCCP preprint IC/92/45 (1992).
10. Davis, M., Summers, J. & Schlegel, D. *Nature* (in the press).
11. Meinhold, P. & Lubin, P. *Astrophys. J.* **370**, L11–L14 (1991).

1. *World Population Prospects* (United Nations, New York, 1988).
2. Keyfitz, N. *Introduction to the Mathematics of Populations* (Addison-Wesley, Reading Massachusetts, 1968).
3. Westin, A. H. *Bioscience* **31**, 523–524 (1981).
4. Leach, D. J. *R. statist. Soc. A* **144**, 94–103 (1981).
5. Koziol, J. A. & Tuckwell, H. C. *N. Z. Statist.* **21**, 35–40 (1986).
6. Levenberg, K. *Q. J. appl. Math.* **2**, 431–444 (1944).
7. Laird, A. K. *Br. J. Cancer* **19**, 278–291 (1965).
8. Tuckwell, H. C. & Koziol, J. A. *Bull. Math. Biol.* **49**, 495–506 (1987).

HTLV-II among pygmies

SIR — Because of their low genetic drift and limited horizontal transmission, human T-cell lymphotropic viruses (HTLV-I and HTLV-II) act as markers of the movement of ancient populations¹. The simian STLV-I isolates do not segregate independently of HTLV-I, suggesting that there has been horizontal transmission between non-human primates and man over long periods. These retroviruses are now being referred to as primate T-cell lymphotropic viruses (PTLVs). PTLVs are presumed to be of Asian origin, as the only indigenous populations with known endemicity of HTLV-II are American Indians, and also because the Asian STLV-I and Melanesian HTLV-I strains are highly divergent². Ancestral PTLV could thus have segregated into a northern Asian branch that was carried over the Bering Strait to America to give rise to HTLV-II, and into an Indo-Malayan branch of PTLV-I to include the current Melanesian and Australian HTLV-I. From there, contacts gave rise to African STLV-I and HTLV-I, perhaps via Madagascar, which then spread in a second wave as the cosmopolitan HTLV-I (ref. 3). A study we carried out in Zaire now questions parts of this hypothesis.

Pygmies are considered to be the oldest inhabitants of central Africa. The Bambuti pygmies are the smallest people

on Earth and the least admixed of pygmies⁴. The Efe are a group of the Bambuti pygmies living in the Ituri Forest of northeastern Zaire. They live in temporary encampments, often in the vicinity of Lese villages. The Lese are people with Bantu physical features but who speak a Sudanic language. Efe pygmies and Lese villagers have always had economic links⁵, but in the rare event that an Efe woman marries a Lese, she and her progeny are no longer part of the pygmy tribe. In 1991, during a survey of physical fitness in Nduye, a village of northern Ituri, we sampled serum from 12 Efe pygmies, all men, and 146 villagers (97 men, 49 women; 120 of them Lese). All 158 adults appeared healthy and none had antibodies against HIV-1 or HIV-2.

Sera reactive in an HTLV-I whole-virus enzyme-linked immunosorbent assay were tested by western blotting against virion HTLV-I and recombinant HTLV-I and HTLV-II specific proteins (see figure). This technique had already distinguished between HTLV-I and HTLV-II in North Americans, Jamaicans and Japanese⁶. In our hands it correctly identified 31 of 32 HTLV-I infections in Zaireans; it failed to type one. It also identified five subjects from Dungu (northeastern Zaire) as HTLV-II positive. Dungu is situated in the savannah north of the Ituri Forest at an aerial distance of 200 km from Nduye. These results are in agreement with another assay that used type-specific synthetic peptides to discriminate be-

tween HTLV-I and HTLV-II (ref. 7).

Antibodies specific against HTLV-II were found in 4 of the 12 male pygmies (95% confidence limit, 9.9–65%) and in 4 of 120 Lese villagers (3 women, 1 man) (95% confidence limit, 0.9–8.3%) (see figure). HTLV-I was identified in one Lese man, in no pygmies, and in one woman from another ethnic group. One HTLV-positive Lese serum could not be typed.

Populations infected with HTLV-II predominantly or exclusively have not been identified before in Africa. The results are best explained by a reservoir of endemic HTLV-II in pygmies. When genomic analysis and virus isolation become feasible in the local situation it will be possible to investigate the relatedness of HTLV-II in pygmies to strains elsewhere. Contact with pygmies may be responsible for the HTLV-II we identified in five people at increased risk of sexually transmitted diseases in Dungu, but not in other areas of Zaire⁷, and may account for the lower prevalence of HTLV-II in the Lese villagers. Three cases of HTLV-II infection have also been reported from Gabon⁸ but its prevalence is low in Africa compared with HTLV-I.

The distinction between HTLV-I as an Old World and HTLV-II as a New World virus does not seem tenable. African pygmies appear to qualify with American Indians as aboriginal populations with endemic HTLV-II. There is no reason for claiming that infection in either population is older than in the other. PTLV sequences from more geographical areas and from other isolated ancient populations are needed to clarify the evolution of PTLVs. A close relative of HTLV-II has not yet been identified in simians, but if there is one it is more likely to be found in Old World primates than in feral New World monkeys, in which there is no evidence so far of either STLV or SIV.

Patrick Goubau

Jan Desmyter

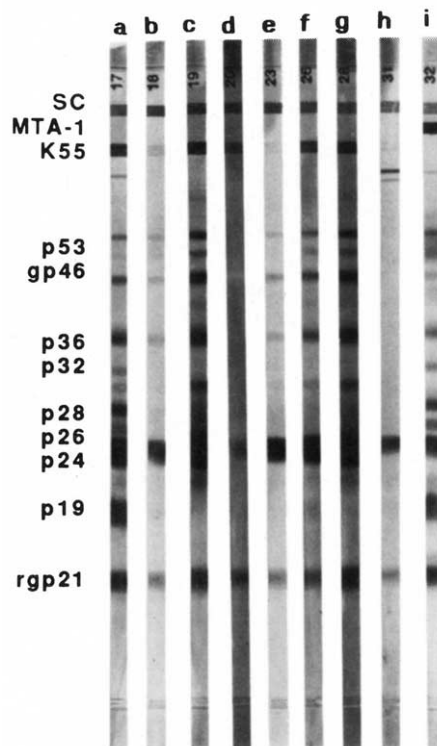
Joseph Ghesquiere

*Rega Institute and University Hospitals,
Katholieke Universiteit Leuven,
Minderbroedersstraat 10,
B-3000 Leuven, Belgium*

Busaka Kasereka

*Hospital of Mambasa, Ituri,
Haut-Zaire, Zaire*

Sera from seropositive Efe pygmies and Lese villagers on HTLV-I western blot strips enriched with recombinant proteins specific for HTLV-I and HTLV-II. S C, serum control; MTA-1, HTLV-I-specific protein; K55, HTLV-II-specific protein; rgp21, recombinant transmembrane protein; native viral proteins (P) are indicated on the left with their respective molecular masses in thousands. Native HTLV-I proteins and rgp21 were separated by SDS-PAGE and blotted onto nitrocellulose; the type-specific MTA-1 and K55 (Diagnostic Biotechnology, Singapore) have been slotted separately. Strips of nitrocellulose were incubated overnight with a 1:100 dilution of serum; after washing and a further incubation with goat anti-human IgG conjugated with alkaline phosphatase, the strips are reacted with 5-bromo-4-chloro-3-indolyl-phosphate and nitroblue tetrazolium (BCIP/NBT). Lanes a–h, HTLV-II antibody-positive sera; lane i, HTLV-I-positive serum, lanes a,d,f,h and i, sera from Lese villagers; lanes b,c,e and g, sera from Efe pygmies. Five of the six HTLV-II antibody-positive sera with nearly complete HTLV-I western blot patterns (lanes a–c and e–g) had a weak or absent P19 band, typically seen with HTLV-II antibody-positive sera. The pattern seen in two HTLV-II-positive Lese sera with an isolated p24 band in the HTLV-I blot (lanes d and h) is seen in many cases of HTLV-II infection (D. Reed, personal communication).



- Gessain, A., Gallo, R. C. & Franchini, G. *J. Virol.* **66**, 2288–2295 (1992).
- Sherman, M. P., Sakseena, N. K., Dube, D. K., Yanagihara, R. & Poiesz, B. J. *J. Virol.* **66**, 2556–2563 (1992).
- Yanagihara, R. 5th Int. Conf. HTLV, (Kumamoto, Japan, May 1992).
- Cavalli-Sforza, L. L. in *African Pygmies* (ed. Cavalli-Sforza, L. L.) 361–426 (Academic, London, 1986).
- Bailey, R. C. & Devore, I. *Am. J. Phys. Anthropol.* **78**, 459–471 (1989).
- Lipka, J. J. *et al. J. Infect. Dis.* **165**, 268–272 (1992).
- Goubau, P. *et al. J. med. Virol.* (in the press).
- Delaporte, E. *et al. Int. J. Cancer* **49**, 373–376 (1991).

Seawater carbon measurement

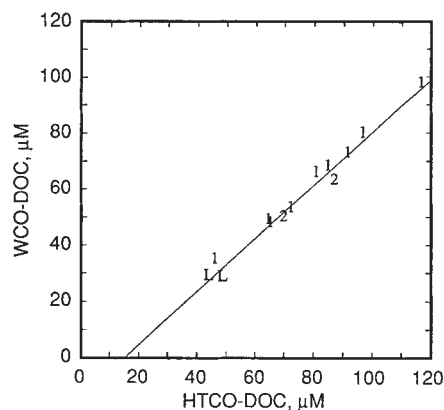
SIR — Ogawa and Ogura¹ compared high-temperature catalytic oxidation (HTCO) and wet chemical oxidation (WCO) methods for measuring dissolved organic carbon (DOC) in sea water. They concluded that both methods gave similar results, although concentrations derived from HTCO were 15–25% higher for bulk ocean water, and 30–40% higher for two low relative molecular mass (less than 1,000) DOC fractions. These results are in contrast to an earlier report² that seawater DOC includes a substantial high molecular mass fraction which can be measured only by HTCO methods and is present in concentrations sufficient to more than double the size of this already immense organic carbon pool³. On analysis of Ogawa and Ogura's data, we find evidence that their HTCO blank could largely explain much of the difference between the reported HTCO and WCO measurements as well as their evidence for an apparent WCO-resistant low molecular mass fraction.

The experimental design used by Ogawa and Ogura compared HTCO and WCO measurements of unconcentrated DOC in bulk ocean water and low molecular mass (> 1k) fractions as well as of concentrated high molecular mass (>10K and >1K) DOC in ultrafiltered samples. The DOC values for the latter samples were reported in terms of initial concentrations after dividing the measured final value by the concentration factor. The authors measured total blanks for both procedures using double-distilled water, obtaining means of 10.5 and 3.8 μM for HTCO and WCO, respectively. They did not correct their reported DOC values for these backgrounds, in part because it was difficult to determine whether the blank signals were derived from the instruments or the distilled water.

A correlation plot of data from Ogawa and Ogura's Tables 1 and 2 shows that the HTCO- and WCO-based measurements of DOC in their unconcentrated bulk seawater and low molecular mass fractions correlate highly ($r^2=0.979$, $n=13$). The best-fit line, however, intercepts the HTCO axis at $15.0 \pm 3.2 \mu\text{M}$ (see figure). This intercept indicates that the HTCO method had an average blank approximately 15 μM higher than for the WCO method, a somewhat greater difference than indicated by the distilled water blanks. Blanks of this magnitude and greater have been observed for various commercial HTCO analysers⁴. It is significant that if the two analyses had equal blanks, but the HTCO instrument

measured a constant fraction of the total DOC not determinable by WCO, the correlation line would extrapolate to the origin.

The larger HTCO blank most affects analyses of more dilute bulk and <1K samples for which the 15 μM higher background accounts for 12–34% of the total measured values. Well over half of the differences in the raw WCO/HTCO ratios listed by Ogawa and Ogura in their Table 1 apparently



Correlation between wet chemical oxidation (WCO) and high temperature catalytic oxidation (HTCO) measurements¹ of unconcentrated dissolved organic carbon (DOC) in sea waters of the western North Pacific. Data for bulk DOC from Tables 1 and 2 of ref. 1 are indicated by these numbers and L indicates concentrations measured for the low-molecular-mass (< 1K) fractions in Table 2 of ref. 1. The plotted line is the best least-squares fit to the data.

result from the higher HTCO blank, as opposed to a greater response of the HTCO versus the WCO instrument. In contrast, the HTCO blank is relatively inconsequential in measurements of the DOC content of the high molecular mass samples, most of which were concentrated to the extent (usually 20 times) that the offset in the HTCO blank corresponds to less than 2% of the total signal.

This interpretation strengthens Ogawa and Ogura's principal finding that the HTCO and WCO methods they compared measure similar concentrations of DOC for various sea waters. Evidence for a substantial fraction of WCO-resistant DOC in the low molecular mass fraction of these water samples, however, is correspondingly weakened. Our re-evaluation demonstrates that instrument blanks and data treatments must be carefully considered when measuring and interpreting DOC con-

centrations. Wide availability and use of DOC-free water would make determinations of DOC analyser blanks much easier.

John I. Hedges

Brian A. Bergamaschi

*School of Oceanography WB-10,
University of Washington,
Seattle,
Washington 98195, USA*

OGAWA AND OGURA REPLY — We agree with Hedges and Bergamaschi that the differences between the measurements could be smaller if the HTCO results are corrected for the system blank. We think that uncertainty in measuring DOC in sea water results mainly from the differences in the scale of blanks with the HTCO instruments and how the results are corrected for the blanks.

Additionally, we wish to emphasize that both low and constant blanks stably obtained throughout the sample measurements are important when the results are corrected for the blanks. As indicated by Hedges and Bergamaschi, the blank values for our HTCO instrument (Shimadzu TOC-5000) seem to be lower than those reported elsewhere, but we do not believe that our instrument is the only type for which this is the case.

We found that elevated blanks could be obtained from deionized/distilled water and that they could be reduced and fixed to some extent by repeated injections of the deionized/distilled water. We conducted the DOC measurements for standard and sample solutions when we confirmed the low and constant blanks after this conditioning process. As shown in Hedges and Bergamaschi's figure, a strong association with an intercept given between the two methods supports the consistency of system blanks for our HTCO analysis.

H. Ogawa

N. Ogura

*Department of Environmental Science
and Conservation,
Faculty of Agriculture,
Tokyo University of Agriculture
and Technology,
3-5-8 Saiwaicho, Fuchu-shi,
Tokyo 183, Japan*

Scientific Correspondence

Scientific Correspondence is a relatively informal section of *Nature* in which matters of general scientific interest, not necessarily those arising from papers appearing in *Nature*, are published. Because there is space to print only a small proportion of the letters received, priority is usually given according to general interest and topicality, to contributions of fewer than 500 words, and to contributions using simple language.

1. Ogawa, H. & Ogura, N. *Nature* **356**, 696–698 (1992).
2. Sugimura, Y. & Suzuki, Y. *Mar. Chem.* **24**, 105–131 (1988).
3. Toggweiler, J. R. *Nature* **356**, 665–666 (1992).
4. Benner, R. & Strom, M. *Mar. Chem.* (in the press).

Philosophical transactions

Jacqueline Reynolds

A Skeptical Biochemist. By Joseph S. Fruton. *Harvard University Press*: 1992. Pp. 330. \$29.95, £23.95.

THERE was a time when science and philosophy were essentially the same discipline. The world produced natural philosophers — not chemists, zoologists or botanists. As knowledge and understanding of nature expanded, scientists and philosophers drifted apart and the former fragmented into different disciplines and subdisciplines. Also in that not too distant past — before the advent of the professional historian of science — it was not uncommon for the ‘man at the bench’ to write his own account of the history of his field. Today, however, most working scientists display little interest in philosophical discussions about the scientific method or the role of hypotheses in discovery, nor do they appear particularly knowledgeable about past events (‘past’ referring to anything before their PhD theses). We now have a bevy of historians of science chronicling and interpreting the various facts of discovery, investigating sociological attitudes or looking at the role of religious belief in the formulation of hypotheses. There appears to be little dialogue and perhaps even a marked antipathy between those historians and philosophers who write about science and those who actually do the work of discovery.

Joseph Fruton, a “skeptical biochemist” and a working scientist, deliberates on the separation of these disciplines. His discourse ranges from a consideration of the scientific method to historical accounts of the development of biology through the past two centuries to advice on how the history of science should be studied. Thrown in for good measure are musings on scientific language and communication. Given the breadth of topics and the candour of the author, this book is guaranteed to raise a few hackles.

Fruton’s encapsulated survey of the ideas and influence of philosophers of science comes from the perspective of an experimentalist, questioning whether present-day philosophers value “thinkers above workers, intellect above craft” — suggesting that much could be learned about the development of the scientific method from a study of the practical manuals used by generations of scientists. A lengthy account of Sanger’s work on the primary structure of insulin is presented as a case history supporting

the triumph of experiment over philosophical theorizing — a reminder to those who write about ‘scientific method’ that “the place of organic chemical practice in the rise and fall of competing theories” should not be neglected.

When he turns to the history of biology since 1800 (“The interplay of biology and chemistry”), Fruton is the trained

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Joseph Fruton: apologist for the working scientist.

organic chemist, presenting an admirable discussion replete with quotations, references and personal views. There is a little of everything here, from embryology to immunology, from cytology to energy-rich phosphate bonds, not to mention extensive philosophical musings on such subjects as the search for unity in nature and its component parts. Not all scientists will agree with Fruton’s view of contemporary scientific achievements. Physical chemists, for example, may be chagrined to find a discussion of forces governing protein structure that relegates hydrophobicity to a relatively minor (and inaccurate) position, with no references to twentieth-century scientists who were prominent in developing this concept. Indeed, we are given a strong impression that physical chemistry (with the exception of X-ray crystallography) has not played a particularly constructive role in the recent history of biological sciences. Perhaps this is because physical chemists tend to look for simplicity and regularity, two vices that we are

told “afflicted some protein chemists before 1940”.

Fruton’s approach to the study of the history of science will strike a responsive chord in many working scientists. “I also believe that a major purpose of professional historians of the various modern sciences ought to be, through meticulous scholarship, the enlightenment of the practitioners in these sciences. If members of the present generation of historians continues [*sic*] to reject this aim ... they will do a disservice both to the advancement of their profession and to the education of future scientists who perhaps may later be inclined to become patrons of their enterprise.” (Amen!) Fruton also takes issue with the tendency of many historians to discount “internal history” (autobiographies, laboratory notebooks, a scientist’s personal account of research developments and so on). Surely there can be little argument that what scientists say about themselves and their disciplines and what they put into the research papers are prime historical data, often giving considerable insight into the character and methods of the individual. An example of the informative nature of internal history can be found here in Fruton’s discourse on “Origins of molecular biology”, which tells the reader as much about the author, his concerns and his biases, as it does about the emergence of this now highly active field.

Fruton concludes with a fascinating discussion of the evolution of scientific language and its effect on communication, both among scientists themselves and more importantly between scientists and those writing about the enterprise. He reminds us of the reasons for adopting what is often described as technical jargon (for example, the need for a single word or symbol that precisely describes an object or an action) and also warns us about the changing meanings of this language as a particular research area develops. Scientists, historians and journal editors could all profit from reading this section.

Fruton has produced a book that demonstrates the sources of tension between scientists on the one hand and historians and philosophers of science on the other, and, as such, it can be commended to anyone engaged in these disciplines. Many of his opinions will be disputed, and so much the better. More dialogue and less tribal exclusivity might profit both the working scientist and those who write about the scientific endeavour. □

Jacqueline Reynolds is at *Tarlswood, Back Lane, Easingwold, North Yorkshire YO6 3BG, UK.*

New in paperback

- *Schrodinger: Life and Thought* by Walter Moore. Cambridge University Press, £10.95, \$19.95. Reviewed by Peter Atkins in *Nature* **341**, 193 (1989).
- *The First Global Revolution: A Report by The Council of The Club of Rome* by Alexander King and Bertrand Schneider. Simon and Schuster, £4.99. Reviewed in the leading article in *Nature* **353**, 95 (1991).
- *The Story of Physics* by Lloyd Motz and Jefferson Weaver. Avon Books, \$12.50.
- *The Ecology of Intercropping* by John Vandermeer. Cambridge University Press, £16.95, \$27.95.
- *The New Physics* edited by Paul Davis, a widely acclaimed survey of developments throughout the whole of modern physics by such contributors as Stephen Hawking, Alan Guth, Anthony Leggett, Chris Isham, Malcolm Longair and Abdus Salam. Cambridge University Press, £16.95, \$27.95. Reviewed by John Mulvey in *Nature* **339**, 350 (1989).

New editions

- *Space, Time and Gravity: The Theory of the Big Bang and Black Holes* (2nd edn) by Robert M. Wald. University of Chicago Press, \$24.95, £19.95 (hbk); \$10.95, £8.95 (pbk).
- *Accretion Power in Astrophysics* (2nd edn) by J. Frank, A. King and D. Raine. Cambridge University Press, £50, \$79.95 (hbk); £19.95, \$37.95 (pbk).
- *Probability and Random Processes* (2nd edn) by G. Grimmett and D. R. Stirzaker. Oxford University Press, £22.50 (pbk).
- *Physics: Structure and Meaning* by Leon N. Cooper. University Press of New England, \$35 (pbk).
- *Introduction to Bacteria for Students of Biology, Biotechnology and Medicine* (2nd edn) by Paul Singleton. Wiley, £11.95, \$26.95 (pbk).
- *Immunology at a Glance* (5th edn) by J. H. L. Playfair. Blackwell Scientific, £9.95 (pbk).
- *Herbivores: Their Interactions with Secondary Plant Metabolites. Volume II, Ecological and Evolutionary Processes* (2nd edn) edited by G. A. Rosenthal and M. R. Berenbaum. Academic, \$99 (hbk).
- *Ethylene in Plant Biology* (2nd edn) by F. B. Abeles, P. W. Morgan and M. E. Salteit Jr. Academic, \$65 (hbk).
- *How to Write and Present Technical Information* by Charles H. Sides. Cambridge University Press, £9.95.

New Journals issue

Nature's New Journals supplement, in which more than 50 publications will be assessed, will appear in the issue of 1 October.

Growth without development

Brian Goodwin

Molds, Molecules, and Metazoa: Growing Points in Evolutionary Biology.

Edited by Peter R. Grant and Henry S. Horn. Princeton University Press: 1992. Pp. 181. \$32.50, £23.50.

DARWIN'S principles of evolution are often placed in the category of enduring scientific theories along with Newton's principles of motion, despite their basic differences — the former are historical, accidental, opportunistic, the latter universal, law-governed, rational. But whereas Newtonian assumptions have undergone transformation with the growth of the subject, Darwin's have not. This volume shows why: the authors examine highly interesting developments in a range of subjects cognate to evolution and, despite significant shifts of emphasis, settle for evolutionary business as usual. The judgement is appropriate, for the book arises from a symposium celebrating the career of John Tyler Bonner, a fervent believer in the explanatory power of natural selection and an eloquent, comprehensive writer on the subject. The title of his, the first essay, says it all: "Evolution and the Rest of Biology".

The six other essays in the volume extend from palaeontology to molecular biology. James W. Valentine addresses the extraordinary phenomenon of the Cambrian explosion, relating diversification of body plans to Bonner's suggestion that the evolution of complexity involves the development of differentiated cell types and the establishment of a hierarchy of pattern-forming genes. But what is being explained? Isn't complexity just being redescribed at other levels? Valentine presents valuable observations on clade extinctions during the Cambrian and concludes that "the clades with the most extinction resistance tend to become increasingly important through time". This is again a restatement of the observations and a tautology. But he is looking for inherent clade properties (not adaptations) that could account logically for the survival of certain taxa rather than others, to escape from the total lottery described by Stephen Jay Gould in *Wonderful Life* (Radius/Norton, 1991). Valentine's concerns are relevant to our current environmental crises and species extinctions, as he points out.

Graham Bell wants to develop a theory of the environment, and he has some intriguing results to inform such a theory: "most of the variance we see in

natural populations, and in particular most of the variance in fitness, is associated with differences among environments rather than differences among genotypes." So "ecology is much more important than genetics". Using norms of reaction as the measure, he shows that organisms respond inconsistently to the environment (that is, there is no optimal fitness), the environment being fractal (complex and self-similar) on all scales. Also, organisms change the environment, resulting in cyclical or chaotic fluctuations in population or community composition. This certainly takes evolution in the direction of a dynamic theory driven by the integrated forces of biogeochemical change. Gaia beckons, but is not acknowledged.

Putting development back into evolution is the theme of the chapters by Mary Jane West-Eberhard and by Leo W. Buss and Matthew Dick, the first in the context of behaviour, the second from the perspective of developmental biology. West-Eberhard sees epigenetic landscapes with their diverging valleys leading to different states of morphology and behaviour as fertile grounds for evolutionary diversification and the source of the polymorphisms that abound in natural populations. Like Bell, she recognizes that natural selection does not result in single-peak adaptive states, but the diversity she emphasizes is within individuals, not in populations. Organisms are diversity-generators that do not have to forgo one state for a fitter one. She illustrates these ideas with copious examples, particularly from the behaviour of social insects.

Buss and Dick develop a similar theme, moving away from the historical perspectives of Darwinism and towards more generative analyses of morphologies and life cycles: "we have no longer any use for the 'Ahnengalleries' (ancestor portrait galleries) of phylogeny . . . We are no better off knowing that we have eyes because our ancestors had eyes", a quote they use from C. O. Whitman (1895). However, they invite us to visit the Ahnengalleries of genes involved in segmentation of *Drosophila*. But the epigenetic map from genes to segments is still missing.

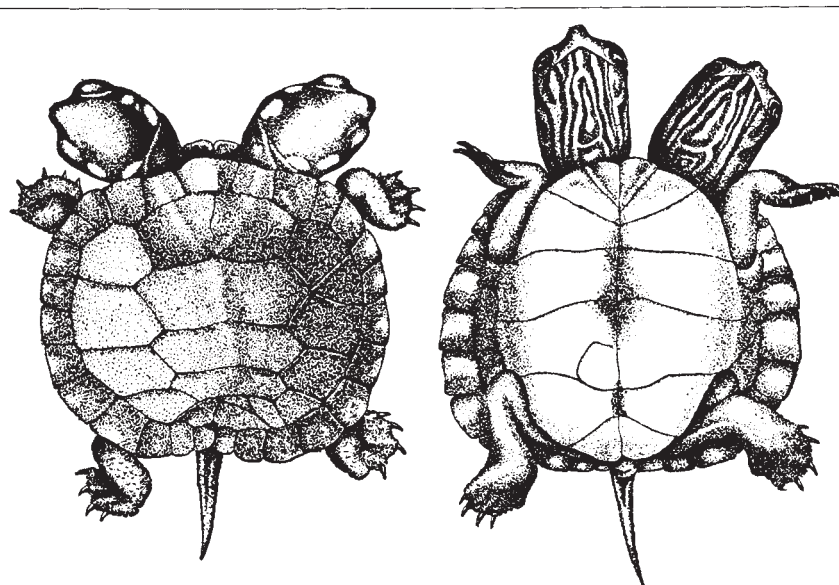
Their enquiry comes closer to this generative level of analysis in an interesting comparative study of different levels of spatial patterning in annelids, akin to the hierarchical levels revealed by segmentation gene domains in *Drosophila*, but operating in asexual reproduction in the adult. Then comes a fascinating description of life cycles, showing how robust and unitary seemingly fragile concatenations of life stages can be. "It would appear that selection at the level of the individual is rather

permissive," they conclude. Development beckons, but is kept in her traditional place.

Mark Kirschner carries the enquiry to the cellular level, where we encounter an interesting paradox. The basic structures and processes of the eukaryotic cell are remarkably conservative, yet the Cambrian explosion occurred and a number of crucial innovations have arisen, making possible important new evolutionary developments. He points out that many modifications are due to different ways of organizing basic cellular 'hardware'. Components of the cell cycle, the cytoskeleton and cell membranes (as in myelin-sheath formation), for example, are altered by changes in genetic 'software', the controlling elements acting through the complex, sloppy, but highly robust and adaptable molecular machinery of enhancers. There is a particularly interesting parallel that emerges between this low-specificity, long-range, dynamic and easily modifiable transcriptional control apparatus of eukaryotes and the equally nonspecific, sloppy, long-range, dynamic and highly adaptable cytoskeletal organization that underlies cell morphogenesis and movement. These make possible the expression of a wide range of cell-differentiation states that contribute to the diversity of metazoan morphology.

Finally, Martin Kreitman warns against any premature euphoria about the power of molecular, let alone developmental, biology to illuminate evolutionary problems. He recalls his first reading, as a graduate student, of Bonner's book *On Development* (Harvard University Press, 1977), in which Bonner attempted to unite evolutionary biology with the molecular principles of development. Kreitman failed to grasp the union, and felt he had somehow missed the point. On re-reading the book in preparation for writing this essay, he discovered that his first impression was correct: the book is divided into two equal but separate parts, one on evolution, the other on development. This is a common fate of writers attempting to correct the imbalance of the modern synthesis, for reasons that are closely connected with Kreitman's own position. He considers that evolutionary biology has a coherent conceptual structure and massive empirical support, as does Bonner. Kreitman foresees its maintenance job being handed over to philosophers and historians. Given such belief in the conceptual completeness of neo-Darwinism, it is perhaps not surprising that evolutionary biology continues to grow without development. □

Brian Goodwin is in the Department of Biology, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK.



Out of print for nearly a century, *Materials for the Study of Variation Treated With Especial Regard to Discontinuity in the Origin of Species* by William Bateson is now reissued with a new introduction by Peter J. Bowler and an essay by Gerry Webster. Published by Johns Hopkins University Press, \$19 (pbk), \$49 (hbk). The picture shows a two-headed specimen of *Chrysemys picta*, 2 or 3 days old (from Barbour, 1888).

Weeding out the jargon

Peter D. Moore

Diversity and Evolution of Land Plants.

By Martin Ingrouille. Chapman and Hall: 1992. Pp. 340. £17.95, \$39.95 (pbk).

PROSPECTIVE biology students at King's College were once asked which aspects of the subject they found most interesting and which the most boring. Plant anatomy and morphology were deemed by far the most boring. Two reasons for this strike me as fairly obvious: the delight that botanists take in obscure jargon — why use the term leaf stalk when you can speak of a petiole? — and the tendency to be satisfied with description for its own sake. Relatively little effort has been made to develop a functional approach in introductory texts. At a time when a lack of well-trained taxonomists is suddenly becoming apparent, it is important that young biologists should be stimulated by the study of structure, adaptation and diversity in plants and animals, and Martin Ingrouille has made a brave attempt to write a text that will capture young minds.

Ingrouille begins by outlining the taxonomic system of nomenclature adopted by botanists and discusses plant structure and the variation displayed by morphological characters as a response to environment. He then considers the evolution of land plants on the basis of fossil evidence, following this with a more

detailed account of plant reproduction, including alternation of generations, the evolution of the angiosperms and the diversification of pollination and breeding systems. A further series of chapters deals with selected topics of botanical interest, such as trees and their architecture, physiological problems and their structural solutions, reproductive strategies, herbivory and, finally, the exploitation of plants through domestication. An extensive glossary and bibliography is appended.

The topics have been well chosen and reflect an up-to-date approach to the study of plant diversity. But does the text avoid excessive jargon and adopt a functional perspective?

Unfortunately, the 'naming of parts' remains a dominant theme, perhaps even an end in itself. As early as page 9 we are supplied with a complex and detailed diagram giving 71 technical terms for leaf shapes, margins and surfaces, including some that are very unlikely to be needed by the trainee botanist (runcinate, lyrate, setose, strigose and farinaceous). The author could easily have provided a diagram to impress the student with the range of leaf form without inducing verbal indigestion at such an early stage. This emphasis on getting the terms right penetrates the entire text; even the fascinating subject of the branching architecture of tree canopies is made unnecessarily heavy and cumbersome by the introduction of

a plethora of terms. Of course, such technical terms become valuable as a shorthand in accurate communication at an advanced stage, but at an introductory level they serve only to erode interest.

Ingrouille has made great efforts to select and arrange material in a way that will stimulate the reader. Leaf arrangements, defence against herbivory (both physical and chemical), mycorrhizae, epiphytic plants and many other topics are given extensive treatment. There are times when more explanatory material would have been helpful, such as in the mechanisms of crassulacean acid metabolism and C4 photosynthesis, *r* and *K* selection (which are neither defined nor explained) and in the use of terms such as 'guerrilla' and 'phalanx' growth forms in vegetative extension of plants. The functional approach, however, is partially adopted, although it is not fully developed until rather late in the text.

My final complaint concerns the diagrams. Although the line drawings are

individually clear and helpful, they are so crowded and complicated by detailed labels and terms that their clarity is often lost. Perhaps the publishers are to blame here, for the text format is open and attractive with wide margins, yet the figures are unacceptably dense and obscure.

Ingrouille has attempted an extremely difficult, perhaps impossible, task, and has in certain respects succeeded. I trust that the book will survive into a second edition, in which some of its present deficiencies could be ironed out. □

Peter D. Moore is in the Division of Life Sciences, Kings College, Campden Hill Road, London W8 7AH, UK.

■ Of related interest is the newly published *Green Plants: Their Origin and Diversity* by Peter R. Bell, a thoroughly revised edition of the earlier *Diversity of Green Plants* by P. R. Bell and C. L. F. Woodcock (3rd edn, 1983). Cambridge University Press, £16.95 (pbk).

Teaching programs

John A. Campbell

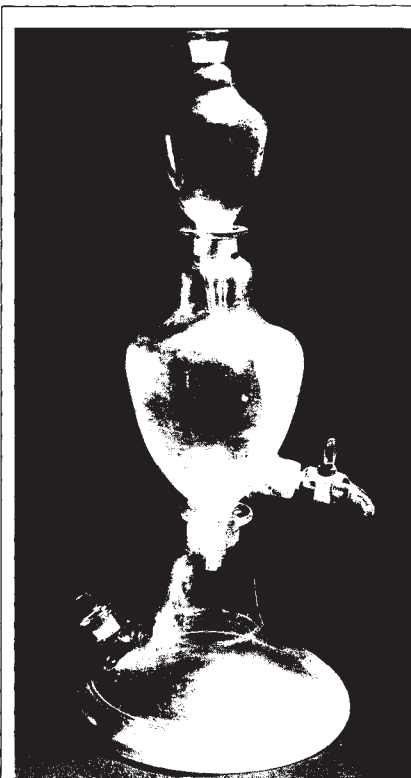
Foundations of Computer Science. By Alfred V. Aho and Jeffrey D. Ullman. Computer Science Press/W. H. Freeman: 1992. Pp. 765. £42.95, \$56.95.

COMPUTER science may well attract more money and attention at present than its academic neighbours in universities, but it is not yet as confident of its identity. Debates on "What should we teach?" still turn up fairly regularly. They take their most solemn form in the United States, particularly under the sponsorship of the Association for Computing Machinery. At least one of the curriculum reports to emerge from these debates looked like a giant inflated pudding, full of facts and more facts, with obvious risks to students' digestions. In a reaction against this dinosaur style, a counter-report published in 1989 argued for three concepts — abstraction, design and theory of computation — as unifying factors for a newer and leaner approach. Although the argument itself is not new, its latest manifestation has been the most successful: dinosaur puddings are now somewhat out of fashion.

Aho and Ullman have a deserved reputation for producing books that work well in particular fields of computer science, such as algorithms and compilers, as both authoritative reference sources and course texts. Here they have taken up the 1989 framework just mentioned and have expanded it into a book of a similar kind, dinosaur-like mainly in

its length, its cover illustration of ponderous grey animals and its weight of 1.78 kilogrammes. The emphasis is on establishing a mathematical foundation for the subject and on treating those parts of it that benefit most from this approach: analysis of algorithms, data models and data structures, automata and grammars and their uses, and applications of standard logics. The authors expect that people who use the book with a course will already have seen introductory material on computing and programming; hence they regard the book as being helpful for second-year or later university teaching. But at the same time it has the traditional Aho-and-Ullman value as a reference source for readers in computer science and other sciences who need information of this type for professional reasons and who are used to handling heavy books.

The three aspects of the framework are not treated equally. Abstraction comes off best, as a distinctive and widely useful feature of computer science. 'Theory' here is primarily the mixture of abstraction and mathematics that has to occur in examination of data structures, algorithms, parsers and compilers, and simple digital circuits. The authors discuss all these applications. However, computability and the less 'applied' parts of the theory of computation receive relatively brief mentions. Lessons about the place of design in computer science are fairly frequent, even though they are more implicit than explicit, and undervalue the aspects of good design that are not reducible to a matter



Eighteenth-century fizz — John Mervin Nooth's improved version of Joseph Priestley's apparatus (1772) for dissolving carbon dioxide gas in water. The gas was produced in the lowest chamber by reacting marble chips with sulphuric acid; water into which the gas dissolved was contained above a valve in the middle vessel, and the bulb at the top provided a head of water to keep the gas under slight pressure. Such apparatus was used to produce simulated spa water and could be found in many homes of the period. The picture is taken from *Science Preserved*, a directory of scientific instruments from collections in the United Kingdom and Else, compiled by M. Holbrook. Published by HMSO, price £35.

of mathematical-technical sophistication. These are observations, not criticisms; even in 765 pages, one cannot do everything.

To sum up, this is a valuable 'how to do it' book that explores much ground in the centre of the present territory of computer science. Its strength is in the technical details; coverage of 'how to appreciate it' and of issues (such as the relation of classical models of computation to alternatives like neurocomputing and subsymbolic computation) that are most likely to change the map of this territory in the foreseeable future is not part of its brief. In its own chosen region, it will probably be equalled by other books, but is unlikely to be bettered. □

John A. Campbell is in the Department of Computer Science, University College, London, Gower Street, London WC1E 6BT, UK.

The formation of binary and multiple star systems

Simon Chapman, Helen Pongracic, Michael Disney, Alistair Nelson, Jacob Turner & Anthony Whitworth

Department of Physics and Astronomy, University of Wales, Cardiff CF2 3YB, UK

Numerical simulations of protostars condensing out of turbulent interstellar gas clouds suggest that binary and multiple protostar systems are common. Such systems form from gas which has been strongly compressed and cooled in radiative shocks. Individual protostars result either from thermal fragmentation of extended layers, or from rotational fragmentation induced by accretion; they are much denser than their surroundings, well separated, and therefore likely to survive as separate entities.

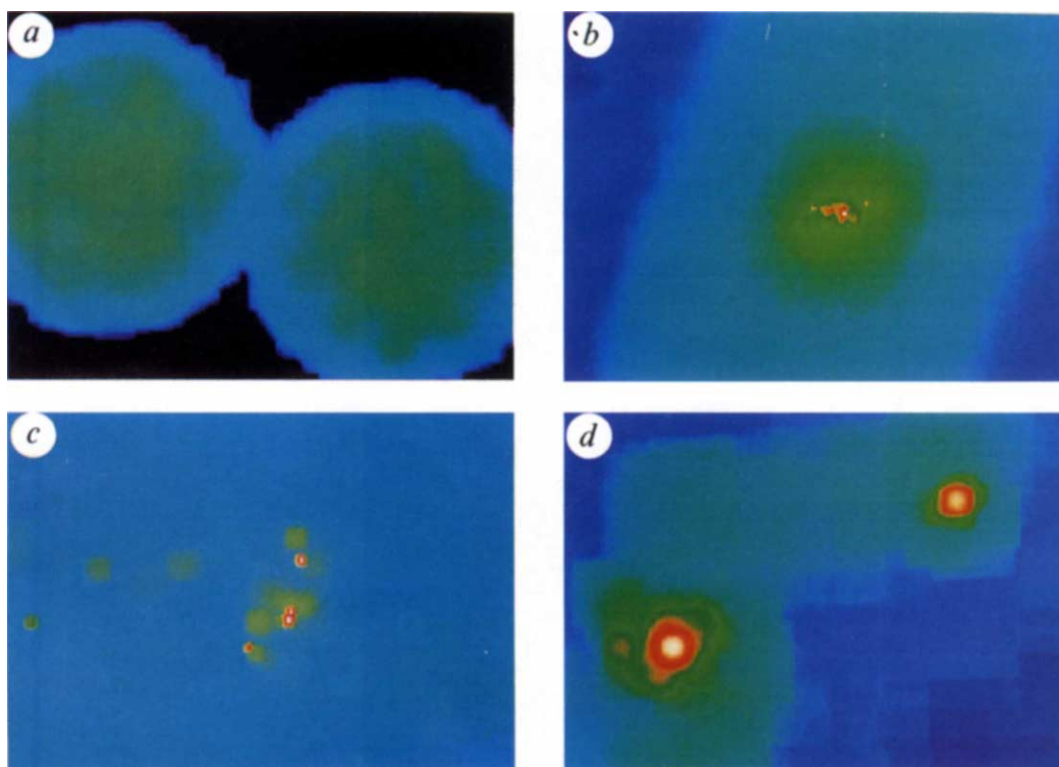
MOST stars are observed to have at least one companion¹, so the question of how individual stars form is inseparable from the question of how binary and multiple stars form². There is much observational evidence indicating that stars form inside giant molecular clouds (GMCs), and that the gas inside these clouds is highly disordered and in supersonic motion³. Therefore to simulate star formation realistically we need a computer code that is capable of handling disordered, supersonic, self-gravitating gas flows, and of following the dynamics through a large range of density and spatial scale. We have developed such a code⁴ by combining smoothed particles hydrodynamics^{5,6} with tree-code gravity^{7,8}. The code has been comprehensively tested. It can handle highly disordered flows because it is three-dimensional and lagrangian. By incorporating a variable particle size, it can follow density increases of up to a factor of 10^{10} . Shocks are modelled using artificial viscosity.

We envisage GMCs as ensembles of either stable or unstable (transient) subclouds, in motion with respect to each other. The gas in our simulations has a piecewise polytropic equation of state, designed to mimic the gross features expected of real

protostellar gas: below a density of $n_{\text{H}_2} = 300 \text{ cm}^{-3}$, it has temperature $T = 100 \text{ K}$; above $n_{\text{H}_2} = 10^4 \text{ cm}^{-3}$, it has $T = 10 \text{ K}$; and at intermediate densities, the gas cools with increasing density according to $T \propto n_{\text{H}_2}^{-2/3}$. The radiative cooling implicit in this equation of state greatly increases the compression factors in supersonic shocks. At the stage where our simulations are terminated, we calculate that trapping of cooling radiation is only just beginning to become significant, so the assumption of isothermality at high density is still realistic.

Even with the aid of diagrams, it is not easy to describe in simple terms the evolution of the simulations. Because of their complexity, they are more readily visualized using false-colour animations, where the colour represents, for example, the column density. In broad terms, supersonic collisions between the initial subclouds lead to strong radiative shocks which compress the gas to high density. The shocked gas then becomes gravitationally unstable and protostellar condensations separate out. At the centres of these condensations, the gas relaxes to form rotationally supported protostellar discs. Interactions between these discs, and with the gas which continues to accrete

FIG. 1 *a*, Time $t=0$, frame size $\Delta X = 7 \times 10^5 \text{ AU}$ by $\Delta Y = 5 \times 10^5 \text{ AU}$, maximum density $n_{\text{H}_2}^{\text{max}} \approx 300 \text{ cm}^{-3}$; two unbound subclouds about to collide. *b*, Time $t = 2.12 \text{ Myr}$, $(\Delta X, \Delta Y)$ as in *a*, $n_{\text{H}_2}^{\text{max}} \approx 2 \times 10^9 \text{ cm}^{-3}$; the extended layer of dense gas which results from the collision. *c*, Time $t = 2.12 \text{ Myr}$, $\Delta X = 7 \times 10^4 \text{ AU}$, $\Delta Y = 5 \times 10^4 \text{ AU}$, $n_{\text{H}_2}^{\text{max}} \approx 2 \times 10^9 \text{ cm}^{-3}$; close-up showing the fragmentation of the layer. *d*, Time $t = 2.82 \text{ Myr}$, $\Delta X = 1.0 \times 10^4 \text{ AU}$, $\Delta Y = 0.7 \times 10^4 \text{ AU}$, $n_{\text{H}_2}^{\text{max}} \approx 7 \times 10^9 \text{ cm}^{-3}$; close-up of the final binary system. The pixels are a false-colour representation of the column density σ through the computational domain, and each colour in the sequence white/red/orange/green/blue represents a logarithmic column-density interval $\Delta \log_{10} [\sigma] = 1$, with blue the minimum.



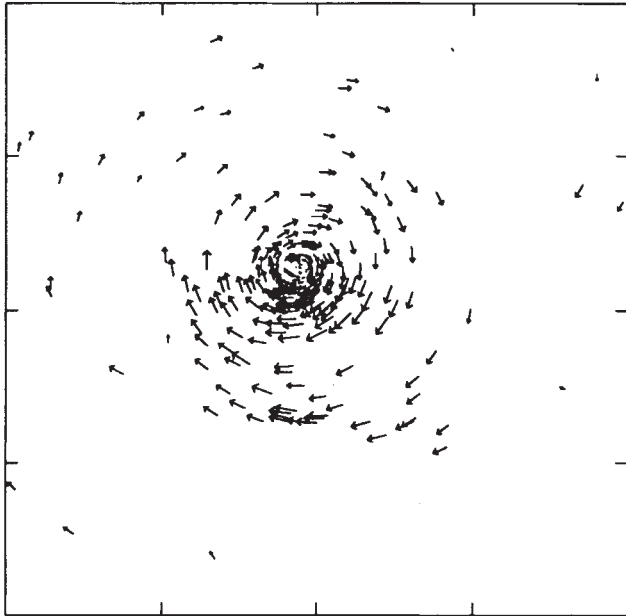


FIG. 2 Velocity vectors for the particles in the larger component of the binary in Fig. 1*d*. The largest velocities at the edge of the disc are $\sim 2.4 \text{ km s}^{-1}$. Frame size, $\Delta X \approx \Delta Y \approx 8 \times 10^3 \text{ AU}$.

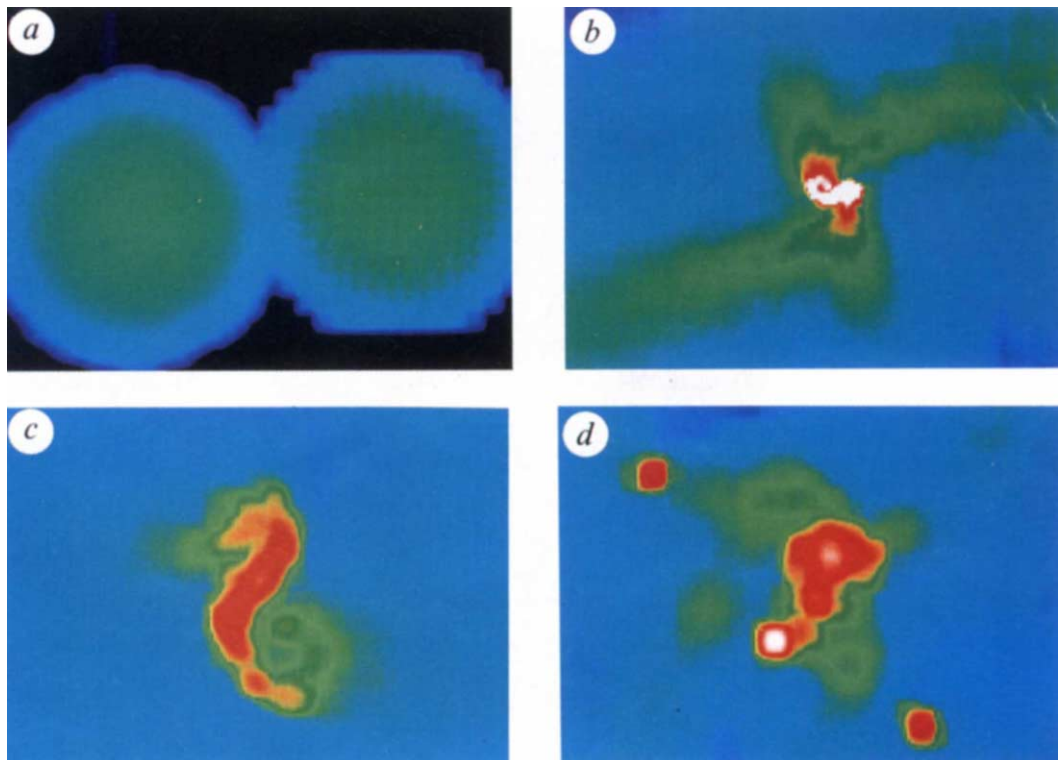
onto them, lead to the formation of stable, multiple systems through a variety of physical processes. The first simulation (Figs 1 and 2) shows several protostellar condensations forming by shock-induced thermal fragmentation, and subsequently merging to form a protobinary. The second simulation (Fig. 3) shows a single protostellar condensation breaking up by accretion-induced rotational fragmentation, and thereby forming a quadruple system, whose eventual fate is uncertain. The third simulation (Fig. 4) leads to two protobinaries in orbit about each other (an hierarchical protobinary); later one of the protobinaries merges and we are left with a stable prototriple.

Our simulations suggest that multiple star formation is a

natural consequence of the highly compressive radiative shocks that result from supersonic turbulence. Because these multiple systems proceed to accrete most of the gas in their neighbourhoods, it is unlikely that their orbits will decay very much because of the drag of the ambient medium. The individual discs have masses in the range 2–40 solar masses ($M_{\odot} = 2 \times 10^{30} \text{ kg}$) and diameters in the range 200–5,000 astronomical units ($1 \text{ AU} = 1.5 \times 10^{11} \text{ m}$).

We give the Mach numbers of shocks relative to the sound speed (0.34 km s^{-1}) of gas that has just entered the cooling portion of the equation of state at the low-density end ($n_{\text{H}_2} \sim 300 \text{ cm}^{-3}$). All our simulations involve $\sim 4,000$ particles, and therefore the fact that we obtain predominantly massive proto-

FIG. 3 *a*, Time $t=0$, $\Delta X = 8 \times 10^5 \text{ AU}$, $\Delta Y = 5 \times 10^5 \text{ AU}$, $n_{\text{H}_2}^{\text{max}} \approx 300 \text{ cm}^{-3}$; two bound subclouds about to collide. *b*, Time $t=0.98 \text{ Myr}$, $\Delta X = 1.6 \times 10^5 \text{ AU}$, $\Delta Y = 1.1 \times 10^5 \text{ AU}$, $n_{\text{H}_2}^{\text{max}} \sim 1.6 \times 10^{11} \text{ cm}^{-3}$; the compact shock and opposing streams of gas that result from the collision. *c*, Time $t=0.98 \text{ Myr}$, $\Delta X = 2 \times 10^4 \text{ AU}$, $\Delta Y = 1.4 \times 10^4 \text{ AU}$, $n_{\text{H}_2}^{\text{max}} \approx 1.6 \times 10^{11} \text{ cm}^{-3}$; close-up showing the moment of fission. *d*, Time $t=1.03 \text{ Myr}$, $(\Delta X, \Delta Y)$ same as *c*, $n_{\text{H}_2}^{\text{max}} \approx 6.3 \times 10^{11} \text{ cm}^{-3}$; close-up showing the multiple system at the end of the simulation. The colour scale is as in Fig. 1.



stars may be a selection effect, as protostars must have $M \geq 2 M_{\odot}$ to be resolved by the code.

Transient subclouds

Figure 1 illustrates the simulated formation of a binary protostar by shock-induced thermal fragmentation followed by merging and cannibalism. Times in megayears (10^6 years), measured from the start of the simulation, are given along with the linear size of the frame (in AU) and the peak density $n_{\text{H}_2}^{\text{max}}$ (in cm^{-3}). Figure 1a shows two subclouds, each with mass $75 M_{\odot}$ and radius 2×10^5 AU, about to collide at relative speed 1.6 km s^{-1} (Mach 4.7) and impact parameter 8×10^4 AU. These subclouds are stable against contraction, but there is no external pressure at their boundaries, so they are transient structures which would disperse if it were not for the fact that they collide first. We presume that many structures in interstellar space are transient in this sense. An extended layer of shocked gas forms when the subclouds collide (Fig. 1b). The large density increase produced by the shock is a result of the high Mach number of the collision and the cooling implicit in the equation of state. The densest part of this layer fragments into several condensations (Fig. 1c). These condensations subsequently interact dynamically and several mergers take place. Figure 1d shows the resulting binary system after it has stabilized and completed two orbits; the scale of Fig. 1d is 70 times smaller than that of Fig. 1a. The components have masses 12.8 and $5.6 M_{\odot}$, diameters 4,000 and 300 AU, and mean separation 11,500 AU. Figure 2 is a close-up of the larger component in Fig. 1d, showing the ordered velocity field of the disc.

The case illustrated in Fig. 1 exemplifies a pattern discernible in many of our simulations. Shock-induced thermal fragmentation (STF) occurs when the dense region produced by the colliding subclouds is an extended layer. Once the layer has sufficient surface density, it fragments in the manner analysed by Larson⁹ to produce an array of protostellar condensations. At the same time these condensations start to fall together on highly radial orbits to form a multiple system near the centre of mass of the layer. The ultimate fate of a particular condensation is usually determined by interactions with other condensations. Energy and angular momentum are redistributed between the condensations. If there is sufficient dissipation, the orbits may shrink considerably, becoming more circular in the pro-

cess¹⁰. Close encounters can lead to merging. Often a smaller condensation is torn apart by the tidal field of a larger condensation and then swallowed up in a violent act of cannibalism.

The extended layers needed for STF result when the colliding subclouds are only slightly centrally condensed, and therefore individually fairly stable against contraction. Consequently there is a significant flux of material over a large front, giving an extended shock. A collision at high Mach number is then required for this shock to produce enough compression to render the gas gravothermally unstable. In addition, a high Mach number ensures that the collision is rapid, so the subclouds do not have time to pinch each other tidally; otherwise tidal pinching can focus the subclouds and thereby reduce the area of the shock. A low impact parameter also increases the extent of the shocked layer by increasing the amount of overlap between the two opposing streams of matter.

Stable subclouds

Figure 3 shows the simulated formation of a protoquadruple by accretion-induced rotational fragmentation. We start with two subclouds, each with mass $75 M_{\odot}$ and radius 2×10^5 AU, about to collide at relative speed 1.6 km s^{-1} (Mach 4.7) and impact parameter 8×10^4 AU (Fig. 3a), as in Fig. 1, but here the subclouds are in detailed hydrostatic balance and contained by a hot rarefied gas. Therefore they are mildly centrally condensed, and intrinsically stable both against contraction and against expansion. As the subclouds approach, they pinch each other tidally, and so the mass flux is focused into a compact shock. After ~ 1 Myr, a rapidly spinning condensation has formed where the two opposing streams of gas impinge on one another (Fig. 3b). This condensation then fragments rotationally (Fig. 3c) to produce a multiple system (Fig. 3d; note that the scale of Fig. 3c and d is 40 times smaller than that of a). The components have masses 32, 12, 3 and $2 M_{\odot}$, diameters 5,000, 800, 400 and 450 AU, and separations between 3,000 and 7,000 AU. Because of limited computer time, we have not followed this simulation far enough to know how the orbits end up.

The case illustrated in Fig. 3 again exemplifies a pattern discernible in many of our simulations. Accretion-induced rotational fragmentation (ARF) occurs when the dense region initially produced by the colliding subclouds is a single,

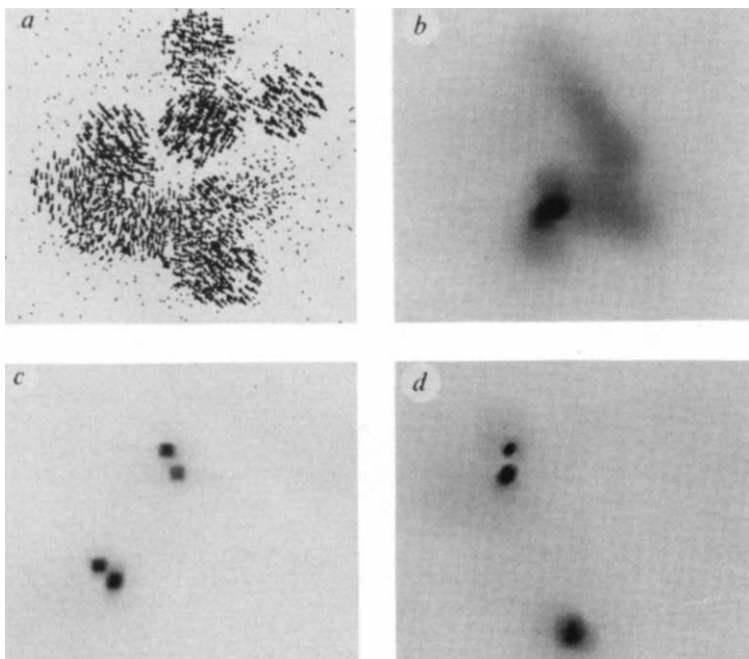


FIG. 4 a, Time $t=0$, $\Delta X=2 \times 10^5$, $\Delta Y=2 \times 10^5$ AU, $n_{\text{H}_2}^{\text{max}} \approx 5 \times 10^3 \text{ cm}^{-3}$: a turbulent ensemble of subclouds. b, Time $t=0.78$ Myr, $(\Delta X, \Delta Y)$ same as a, $n_{\text{H}_2}^{\text{max}} \approx 7 \times 10^{10} \text{ cm}^{-3}$: the condensation resulting from the shock dissipation of the turbulent energy. c, Time $t=0.78$ Myr, $\Delta X=1.2 \times 10^4$ AU, $\Delta Y=1.2 \times 10^4$ AU, $n_{\text{H}_2}^{\text{max}} \approx 3 \times 10^{10} \text{ cm}^{-3}$: close-up showing the hierarchical binary soon after formation. d, Time $t=0.82$ Myr, $(\Delta X, \Delta Y)$ same as c, $n_{\text{H}_2}^{\text{max}} \approx 7 \times 10^{10} \text{ cm}^{-3}$: close-up showing the final stable hierarchical triple. In a we show the velocity vectors of the individual particles, whereas in b, c and d we show a grey-scale representation of column densities through the computational domain. The full range of greys represents a logarithmic column-density interval $\Delta \log_{10} [\sigma] = 5$.

compact, rapidly spinning condensation, which continues to accrete. The detailed dynamics of ARF normally fit a pattern between the following two extremes. At one extreme, a rapidly rotating condensation elongates to form a bar. As the material accreting onto the bar brings in more angular momentum and is therefore directed towards the ends of the bar, the bar cleaves into two protostars. The tidal stresses exerted by these protostars on each other may cause them to fragment further^{11,12}. At the other extreme, a less rapidly rotating condensation repeatedly excites spiral arms which then detach themselves to produce satellite condensations¹³. These satellites grow at the expense of the central condensation by intercepting the continuing accretion flow. The satellites have an advantage over the central condensation because they can convert most of the angular momentum of the accreted material into orbital motion rather than spin, and this is energetically favourable¹⁴. Eventually they may become so large that they tidally disrupt the original central condensation. ARF differs from the rotational fragmentation described, for example, by Boss¹⁵ because in ARF a key role is played by the angular momentum of the continuing accretion flow; in this regard it is more akin to the floccule theory¹⁶.

The compact spinning condensations needed for ARF result when the colliding subclouds are close to gravitational instability, because such subclouds are centrally condensed. Consequently they present each other with a significant flux of material over only a small front, and this leads to a localized shock. Because the subclouds are close to instability, a collision at low Mach number suffices to render the shocked gas gravothermally unstable. A low Mach number also ensures that the approach to collision is slow, so the subclouds have time both to veer towards each other and to pinch each other tidally, thereby focusing the flux of material entering the shock and reducing the area of the shock even further. A high impact parameter increases the spin angular momentum of the initial condensation and the orbital angular momentum of the subsequent accretion flow, thereby promoting ARF. But with large impact parameter the Mach number must not be too high or the clouds hardly interact at all. ARF does not seem to favour components of closely equal mass.

Hierarchy of subclouds

Figure 4 is an example of the complex flows that a code has to handle to simulate collective star formation and facilitate a realistic comparison with observation. Here we have attempted to represent more realistically the internal structure of one of the subclouds involved in the simulations of Figs 1 and 3. Observations suggest that GMCs have a fractal structure¹⁷, with a hierarchy of sub-subclouds within subclouds over a wide range of scales. Figure 4a shows the initial conditions for an ensemble of 20 sub-subclouds, each of mass $4 M_{\odot}$ and radius 0.2 pc,

having mean separation 0.6 pc and moving randomly with a mean bulk speed of 0.59 km s^{-1} (Mach 1.7). The space between the sub-subclouds is occupied by $20 M_{\odot}$ of hot rarefied gas, and the whole ensemble, representing a single subcloud, is roughly virialized. But because collisions between the sub-subclouds are highly inelastic, they quickly dissipate their bulk kinetic energy and merge to form a dense layer (Fig. 4b). This layer then fragments to produce an hierarchical protobinary. Figure 4c shows this part of the computational domain in detail (the scale in Fig. 4c and d is 16.0 times smaller than in a). There are two relatively close binary systems orbiting one another, but one of these close binary systems later merges to form a single protostar. Figure 4d shows the same region at the end of the simulation. By this stage the surviving close binary system has completed six orbits and seems to be stable against merger or disruption; its components have masses 11.2 and $4.9 M_{\odot}$, diameters 600 and 150 AU, and separation 1,200 AU. The third protostar (the result of the merger) has completed one orbit, and it too seems to be stable; it has mass $8.8 M_{\odot}$, diameter 800 AU and an orbit of radius 7,500 AU. Inspection of the animation of this run shows that the wide binary system forms by STF, as does the close binary system that eventually merges, whereas the surviving close binary system forms by ARF. Over 80% of the mass in Fig. 4d is in the three protostars, and so drag is unlikely to cause the orbits to decay much more. We appreciate that the kinetic support of a real subcloud is probably not dissipated on a crossing timescale, as has happened here; but equally, the substructure in real subclouds will mean that collisions between subclouds are effectively many-body collisions between smaller sub-subclouds, and this is the concept we have explored in this simulation.

Discussion

These calculations demonstrate that it is possible to simulate the formation of multiple protostars, starting from realistic dynamical initial conditions. The dimensions of the systems formed are similar to those of observed putative protostellar discs^{18,19} and pre-main-sequence multiple systems^{20,21}. Given the wide range of initial conditions observed in regions of star formation, and the considerable nonlinearity of routes to star formation demonstrated by the simulations reported here, it is clear that a great variety of final configurations can result. Binary and multiple systems seem to be the rule rather than the exception. The individual protostars tend to be massive, but there is no strong tendency for binary components to have nearly equal masses. In principle, with a large ensemble of calculations of this type, we could make quantitative theoretical predictions of how the statistics of binaries (such as primary mass, mass ratio and separation) vary with epoch, metallicity and environment. □

Received 1 June; accepted 6 August 1992.

1. Abt, H. A. *Rev. Astr. Astrophys.* **21**, 343–372 (1983).
2. Pringle, J. E. in *The Physics of Star Formation and Early Stellar Evolution* (eds Lada, C. & Kylafis, N.) 437–448 (Kluwer, Dordrecht, 1991).
3. Falgarone, E. & Phillips, T. G. *Astrophys. J.* **359**, 344–354 (1991).
4. Pongracic, H. et al. *Mon. Not. R. astr. Soc.* **256**, 291–299 (1992).
5. Lucy, L. B. *Astr. J.* **82**, 1013–1024 (1977).
6. Gingold, R. A. & Monaghan, J. J. *Mon. Not. R. astr. Soc.* **181**, 375–389 (1977).
7. Barnes, J. & Hut, P. *Nature* **324**, 446–449 (1986).
8. Hernquist, L. *Astrophys. J. Suppl. Ser.* **64**, 715–734 (1987).
9. Larson, R. B. *Mon. Not. R. astr. Soc.* **214**, 379–398 (1985).
10. Pringle, J. E. *Mon. Not. R. astr. Soc.* **239**, 361–370 (1989).
11. Boss, A. P. *Nature* **351**, 298–300 (1991).
12. de Felice, F. & Sigalotti, L. D. G. *Mon. Not. R. astr. Soc.* **249**, 248–261 (1991).

13. Durisen, R. H., Gingold, R. A., Tohline, J. E. & Boss, A. P. *Astrophys. J.* **305**, 281–308 (1986).
14. Lynden-Bell, D. & Pringle, J. E. *Mon. Not. R. astr. Soc.* **168**, 603–637 (1974).
15. Boss, A. P. *Astrophys. J. Suppl. Ser.* **62**, 519–552 (1986).
16. McCrea, W. H. *Proc. Roy. Soc. A* **256**, 245–266 (1960).
17. Falgarone, E., Phillips, T. G. & Walker, C. K. *Astrophys. J.* **378**, 186–201 (1992).
18. Sandell, G., Aspin, C., Duncan, W. D., Russell, A. P. G. & Robson, E. I. *Astrophys. J.* **376**, L17–L20 (1991).
19. Wooten, A. *Astrophys. J.* **337**, 858–864 (1989).
20. Zinnecker, H., in *Low Mass Star Formation & Pre-Main Sequence Objects* (ed. Reipurth, B.) 447–469 (European Southern Observatory, Garching, 1989).
21. Clarke, C. J. *Nature* **357**, 197–198 (1992).

ACKNOWLEDGEMENTS. We thank G. Allen and P. Fayers for help with graphics, and A. Boss for helpful comments.

Cystic fibrosis in the mouse by targeted insertional mutagenesis

Julia R. Dorin*, Paul Dickinson*, Eric W. F. W. Alton†, Stephen N. Smith†, Duncan M. Geddes†, Barbara J. Stevenson*, Wendy L. Kimber*, Stewart Fleming‡, Alan R. Clarke§, Martin L. Hooper§||, Louise Anderson||, Rosa S. P. Beddington|| & David J. Porteous*¶

* MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

† Ion Transport Laboratory, National Heart and Lung Institute, Manresa Road, London SW3 6LR, UK

‡ Department of Pathology, University of Edinburgh Medical School, Teviot Place, Edinburgh EH8 9AG, UK

§ Cancer Research Campaign Laboratories, Department of Pathology, University of Edinburgh Medical School, Teviot Place, Edinburgh EH8 9AG, UK

|| AFRC Centre for Genome Research, Kings Buildings, West Mains Road, Edinburgh EH9 3JQ, UK

Cystic fibrosis is a fatal genetic disorder which afflicts 50,000 people worldwide. A viable animal model would be invaluable for investigating and combating this disease. The mouse cystic fibrosis transmembrane conductance regulator gene was disrupted in embryonal stem cells using an insertional gene targeting vector. Germ-line chimaeras were derived and the offspring of heterozygous crosses studied. These homozygous mutant mice survive beyond weaning. *In vivo* electrophysiology demonstrates the predicted defect in chloride ion transport in these mice and can distinguish between each genotype. Histological analysis detects important hallmarks of human disease pathology, including abnormalities of the colon, lung and vas deferens. This insertional mouse mutation provides a valid model system for the development and testing of therapies for cystic fibrosis patients.

CYSTIC fibrosis is a fatal autosomal recessive disorder characterized by defective chloride transport in epithelial cells and excess mucus secretion^{1,2}. Meconium ileus, or neonatal intestinal blockage, is virtually diagnostic of the disease, but is only seen in ~10% of cases^{1,2}. Pancreatic insufficiency and malabsorption in the gut are common features. Males are typically infertile owing to blockage or absence of the vas deferens¹⁻³. Chronic opportunistic lung infection, probably due to viscid mucus production, is the principal cause of premature death^{1,2}. There is, however, a broad spectrum of presentations and variable severity of clinical manifestation^{1,2}; mildly affected patients may have only moderately elevated sweat chloride, minimal lung deficiency and pancreatic sufficiency. Improved nutritional care, pancreatic enzyme supplementation, aggressive antibiotic treatment and physiotherapy, and earlier diagnosis have all combined to improve dramatically the life expectancy of cystic fibrosis patients. Nevertheless, this is still essentially a fatal disease and patients infrequently survive beyond the age of 30.

Remarkable progress in analysing the biochemical defect in cystic fibrosis has been made since the positional cloning and primary characterization of the gene⁵⁻⁶. The encoded protein was named the cystic fibrosis transmembrane conductance regulator (CFTR) because of its predicted role and from comparison with other gene products of known structure and function⁵. About 70% of patients were found to carry a single mutation, $\Delta F508$, a three-base pair (bp) deletion which removes a single phenylalanine residue⁶. Subsequently, over 170 independent CFTR mutations have been identified⁷. The $\Delta F508$ mutation is almost invariably associated with severe disease and pancreatic insufficiency⁸, but a small subset of missense mutations is associated with pancreatic sufficiency⁸. Homozygous nonsense mutations have been reported both in patients with very mild^{9,10}, or with severe clinical symptoms^{11,12}, and with or without pancreatic insufficiency. Missense mutations indicate residues likely to be critical for normal function. This supposition is largely borne out by ectopic expression studies¹³.

Mutation detection allows accurate prenatal diagnosis in

many families, but the majority of new cases of this autosomal recessive disorder arise in the absence of a previous family history. In most populations, detection rates still fall short of those required for comprehensive screening⁷. The aim of experimental research must be to improve diagnosis, disease management and therapy. Certain limitations of clinical and *in vitro* studies emphasize the need for an animal model. It is noteworthy that all but one of 60 missense mutations in man occur at amino-acid residues that are conserved in the mouse¹⁴. This argues for a conservation of CFTR function which we test here by a targeted, insertional disruption of the mouse *Cftr* gene in embryonal stem cells¹⁵. The viability of homozygous mutant offspring together with the pathology and electrophysiology of their epithelial-lined tissues suggest that this mutation creates a mouse that closely models important features of cystic fibrosis in man.

Gene targeting and germ-line transmission

We have described elsewhere¹⁵ the construction of the insertional vector pIV3.5H designed to disrupt the CFTR gene in exon 10. Briefly, a 3.5-kilobase (kb) *HindIII* fragment of DNA isogenic with the E14 embryonal stem cell line and which encompasses part of intron 9 and extends into exon 10 was cloned into the pMC1neopolyA vector. This insertional vector was linearized at the unique Asp 718 site, electroporated into the E14 embryonal stem-cell line, and G418-resistant clones screened by Southern blotting for the presence of a novel, diagnostic fragment using a flanking genomic probe, 1.2H, not present in the vector (Fig. 1a, b). Correctly targeted clones were obtained at a frequency of about 2% of G418-resistant clones. Clones retaining a modal chromosome number of 40 and the ability to differentiate *in vitro* were injected into C57 Bi/6 blastocysts and transferred to pseudopregnant female mice to obtain chimaeric offspring. Of 24 offspring, 13 (ten male) were chimaeric, as judged by coat colour, and eight males were mated to MF1 (Olac) females to test for germ-line transmission. Seven males produced offspring of grey coat colour (genotype $+ / p$, c / c^{ch}), indicative of germ-line transmission. These were genotyped by Southern blotting using the 1.2H probe. About

¶ To whom correspondence should be addressed.

half (54 out of 113) were heterozygous for the wild-type 6-kb *Xba*I fragment (wt) and the novel, mutant 5-kb *Xba*I-*Sal*I fragment (m) (Fig. 1b). We have designated the insertional mutation as *Cftr*^{TgH(neoim)1HGU}, but we abbreviate this to *cf*. The *cf*/+ mice (heterozygous for the insertional mutation) were phenotypically normal.

Production of viable homozygous mutant mice

One male (derived from chimaera 42) was mated to three females (all derived from chimaera 48) and produced 24 offspring, of which 8 were wild-type (+/+), 10 were heterozygous (+/*cf*) and 6 were homozygous mutant (*cf*/*cf*). Figure 1b shows an analysis of part of this pedigree. A single mouse died perinatally, was histologically normal, and genotyped as being *cf*/+. All other mice survived beyond weaning and *cf*/*cf* mice showed no overt failure to thrive over the study period (up to 30 days postpartum).

Electrophysiology

If targeted disruption of the *Cftr* gene effectively abolishes gene function, we would expect to see electrophysiological changes consistent with the absence of a functional CFTR chloride channel in all *cf*/*cf* mice, irrespective of disease status. A blind trial was done on nine mice, comprising one control mouse plus eight offspring selected from the three litters in the study group. These eight coded mice, and additional mice, were also examined for pathological changes (see below). Gastrointestinal and respiratory tract potential differences (p.d.s) were determined *in vivo* by an established method¹⁶, modified for the mouse (E.W.F.W.A., S.N.S. and D.M.G., manuscript in preparation). The pooled results are given in Table 1. All four *cf*/*cf* mice were unequivocally identified. Baseline rectal p.d. in the homozygous mutant animals was significantly lower than the combined heterozygous/wild-type group. Colonic/caecal p.d. was similarly significantly reduced in homozygous mutants. After perfusion of each of the above regions with a combination of forskolin (an activator of adenylate cyclase) and zardaverine (a phosphodiesterase inhibitor), in the presence of amiloride (an inhibitor of sodium transport), a significant reduction in the resultant hyperpolarization was seen in all homozygous mutant mice from (−5.5 to −6.7 for rectal and −3.7 to −3.4 for colon compared with non-mutant mice (−16.2 to −22.3 for rectal and −10.4 to −12.0 for colon)). Nasal p.d. was significantly increased

in the homozygous mutant mice, but no difference was discernible throughout the lower airways below the larynx. Perfusion of the trachea with a low-chloride solution showed a significant reduction in the resultant hyperpolarization in each of the *cf*/*cf* animals. Heterozygotes (range −10 to −11 mV) demonstrated rectal p.d.s intermediate between *cf*/*cf* (range −4 to −9 mV) and +/+ (range −18 to −24 mV) mice, as well as either a reduced response to forskolin and zardaverine in the rectum, or to low-chloride perfusion in the trachea, neither of which overlapped with the *cf*/*cf* values. In summary, all nine animals were studied without prior knowledge of genotype and all were correctly identified. The animals, still coded, were subjected to pathological examination.

Pathology

Histological examination was done on formalin-fixed tissue from the three litters, described above. The histology of wild-type and heterozygous mice is indistinguishable and completely normal. Figure 2 shows histologically stained tissue sections of the colon from a mutant *cf*/*cf* (a), and a heterozygous *cf*/+ mouse (b). The mutant *cf*/*cf* mouse shows pathological changes consistent with those seen in cystic fibrosis. There was mild colon dilatation with abnormal mucus accumulation and a resultant distension of epithelial cells, most prominent in the crypts. The gonads were normal, but in one *cf*/*cf* male increased mucin was also seen in the vas deferens (Fig. 2c, d). One mouse, at 30 days postpartum, showed mild focal pulmonary atelectasis (Fig. 2e, f), consistent with presymptomatic lung disease, and mild dilatation of salivary ducts (data not shown). There was no overt histological abnormality in the pancreas and this is consistent with the average body weight gain of the mice (data not shown). In two out of six *cf*/*cf* mice examined, we failed to detect overt histopathological abnormalities, consistent with a variable penetrance, as seen in man. These two mutant mice, however, were readily identified by p.d. measurements (see above).

Discussion

We have provided evidence that mice homozygous for an insertional disruption of the CFTR gene mimic important features of cystic fibrosis, with altered chloride ion transport in all cases, and variable, but diagnostic pathology in four out of six *cf*/*cf* mice examined. In the large intestine, cystic fibrosis patients

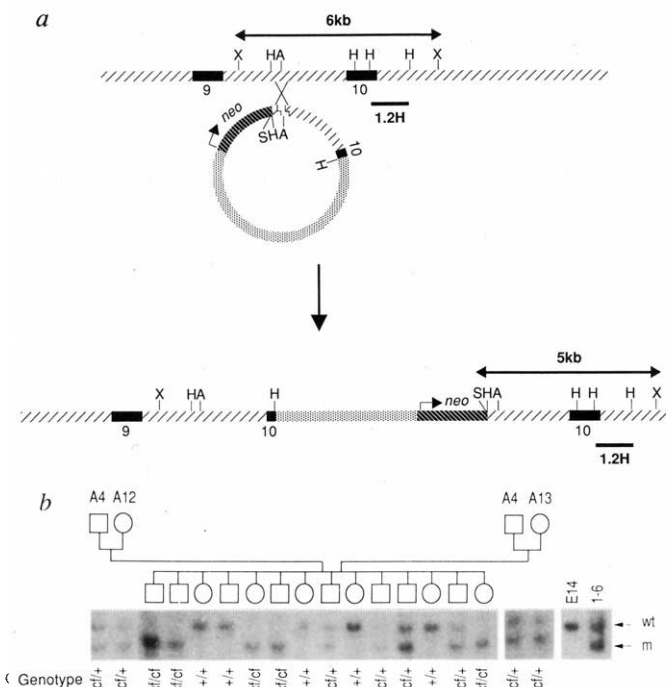


FIG. 1 Insertional disruption of the murine CFTR gene. *a*, Targeting scheme. The design of the insertional vector has been described in detail previously¹⁵. The figure illustrates the genomic structure, the point of insertion in intron 9 and the predicted gene disruption. Probe 1.2H is used to follow insertion events by Southern blot analysis of putative targeted cell lines and derived mice. The diagnostic restriction fragments are indicated. Abbreviations: S, *Sal*I; H, *Hind*III; X, *Xba*I; A, *Asp*718. *b*, Genotype analysis of heterozygous *cf*/+ matings. Segregation and Southern blot analysis of two of the three sibships described in the study are shown. Following the proposed rules of the International Standardized Genetic Nomenclature for Mice, we designate the *Cftr* allele created by the transgene insertion as *Cftr*^{TgH(neoim)1HGU}. Male heterozygote A4 (derived from chimaera 42) was crossed with female heterozygotes A12 and A13 (both derived from chimaera 48). The 14 offspring were genotyped by Southern blot analysis of tail-tip DNA taken at 17 days. DNAs were digested with *Xba*I + *Sal*I, transferred to nylon membranes and probed with the 1.2H genomic probe indicated in *a*. The upper hybridizing fragment of ~6 kb represents the wild-type (wt) fragment, whereas the lower hybridizing fragment of ~5 kb is diagnostic for the insertional mutation (m). Analysis of the E14 embryonic stem-cell line used in the targeting experiment¹⁵, together with the targeted clone 1.6, is shown for comparison.

demonstrate a reduced rectal p.d. (ref. 17), as well as the characteristic physiological defect of impaired chloride secretion, in response to elevation of cyclic AMP¹⁸. We have previously shown that *in vitro* murine caecum responds to forskolin and zardaverine, in the presence of amiloride, with an increase in p.d. and equivalent short circuit current, which can be largely inhibited by bumetanide¹⁹. Therefore, this response is very likely to represent chloride secretion. The large intestines of the four *cf/cf* mice in this study thus seem to demonstrate two of the electrophysiological abnormalities of the gastrointestinal tract of cystic fibrosis patients. An elevated nasal p.d. is one of the hallmarks of cystic fibrosis¹⁶ and is used in the clinical diagnosis of patients. Human nasal epithelium predominantly absorbs sodium in the resting state and the raised p.d. in cystic fibrosis relates to an increase in this absorption²⁰. We have previously shown that murine tracheae display ion transport properties

similar to those characteristic of humans¹⁹. Although technical considerations precluded the use of nasal amiloride in this study, the significantly increased nasal p.d. in the *cf/cf* mice argues strongly for an increased sodium absorption in the nasal mucosa of these animals. The lower airway p.d. of a limited number of cystic fibrosis patients has been shown to be increased¹⁶. This is not consistently observed^{21,22} and was not seen in this study. Nevertheless, the tracheae of all the *cf/cf* animals tested demonstrated the characteristically reduced permeability of these tissues²³. We were able to distinguish heterozygotes on the basis of a minimum of two of the above criteria, although clearly a larger number of animals needs to be studied to confirm these findings.

We have no evidence yet for abnormal pancreatic histology, which is possibly due to the ameliorating effect of the large volume of acinar fluid seen in the mouse but not in man (B. Argent, personal communication). Thus, the ductal blockage that typifies most cystic fibrosis patients might be expected to be less severe in mice, or to develop later. Differences between the pulmonary anatomy of mouse and man has called into question whether a mouse mutant for *Cftr* would mimic the human lung disease⁷. But the one *cf/cf* mutant mouse described here that exhibited clear signs of lung damage is significant and consistent with the age of onset of lung disease in human patients. It will be of interest to monitor the effect of ageing on lung disease, as well as the consequences of exposure to lung pathogens such as *Pseudomonas aeruginosa*. The mucoid form of *P. aeruginosa* shows a particular association with cystic fibrosis and accelerates lung damage¹. The presence of mucin in the vas deferens of one male examined also mimics the pathology in cystic fibrosis²⁴ and thus this mouse model may be valuable for studying male infertility. In summary, the mutant *cf/cf* mice display the well established electrophysiological defect associated with cystic fibrosis and a relatively mild pathology, as judged by histology at the time of examination (10–30 days postpartum). A long-term study of survival and pathology is in progress.

The age of onset and severity of cystic fibrosis is extremely variable. At one extreme, neonates die of meconium ileus, whereas at the other the disease is not diagnosed until late adolescence or beyond²⁵. With the exception of pancreatic status, however, there is as yet no clear correlation between specific mutation and clinical phenotype. Rather, the weight of evidence favours the existence of additional genetic factors influencing the expression of the disease²⁶. In this regard, a mouse model has an additional important experimental advantage in that the genetic background can be carefully controlled and modified. We anticipate that such experiments will clarify the relationship between genotype and phenotype and allow us to address the question of contributory genetic and epigenetic factors. Ashkenazi Jews¹² and Georgian Soviets¹¹ homozygous for 'null' mutations have been reported with severe phenotypes, whereas in populations of mixed ethnic origin comparable mutations were associated with mild disease^{9,10}. The mice described here, which are the result of outcrossing strain 129 (E14 genotype) to an MF1 outbred genetic background, display a mild phenotype. Experiments are in progress to determine the phenotypic effect of crossing the insertional mutation onto various inbred backgrounds.

We chose to adopt an insertional targeting strategy for two practical reasons. Firstly, previous studies had suggested an increase in the ratio of homologous to non-homologous recombination with insertion versus replacement vectors²⁷. Although this seems to be true for the *Cftr* gene (0.05% by replacement²⁸ compared with 2% by insertion¹⁵, using overlapping subfragments of the same genomic clone of isogenic DNA), a recent reappraisal questions the validity of the earlier conclusions²⁹. The second advantage of an insertional strategy is that it can be readily adapted to create specific mutations by the 'hit-and-run' procedure³⁰. By subtle modification of our insertion vector,

TABLE 1 Measurement of potential differences in respiratory and gastrointestinal tract

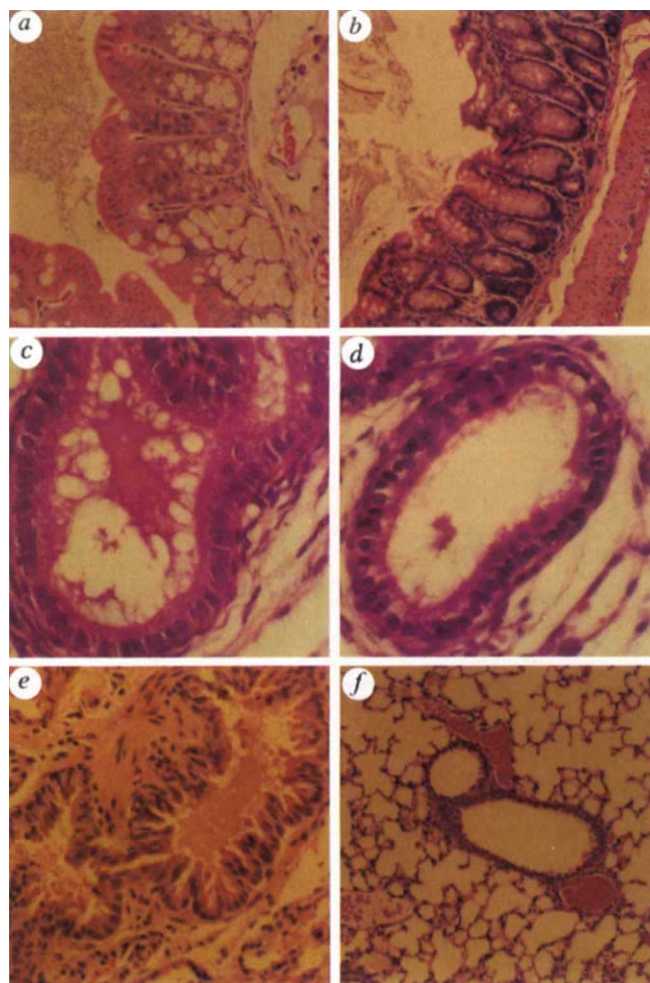
Tissue	<i>cf/cf</i> (n=4)	<i>cf/+</i> , <i>+/+</i> (n=5)
Buccal	-13.8 (±1.4)	-14.2 (±0.7)
Rectal	-5.5 (±1.2)*	-16.2 (±2.6)
Rectal FZ	↑1.2 (±0.4)*	↑6.1 (±1.3)
Caecum/colon baseline	-3.7 (±0.6)**	-10.4 (±0.9)
Caecum/colon FZ	↓0.3 (±0.2)*	↑1.6 (±0.5)
Nasal	-27.3 (±1.9)**	-19.0 (±1.4)
Trachea low Cl ⁻	↑1.0 (±0.6)*	↑4.8 (±0.6)
Upper trachea	-15.8 (±1.7)	-12.3 (±2.3)
Bronchial	-7.0 (±1.9)	-7.5 (±0.6)
Peripheral	-4.3 (±1.9)	-3.5 (±0.6)

Gastrointestinal and respiratory tract potential difference measurements were previously established in a set of MF1 mice (data not shown, results summarized below). For this experiment, baseline measurements were first confirmed on a normal control mouse. Eight experimental mice were sampled from three (*cf/+*) × (*cf/+*) litters and analysed in a blind trial. For statistical analysis of significance, the results from individuals identified by DNA analysis as *cf/cf* were pooled and compared with the combined *cf/+* and *+/+* genotypes. Means and (in brackets) standard errors of the means (s.e.m.) are shown, and significant differences between the two classes are asterisked. An upward-pointing arrow indicates an increase towards more negative values, and conversely for a downward-pointing arrow. Buccal p.d. was used to validate the integrity of the circuit before further measurements. Nasal and airway p.d.s were recorded after a distal tracheostomy. Tracheal p.d. was recorded from the proximal trachea. 'Peripheral airways' represent the most peripheral part of the airways reached by the exploring electrode of 0.5 mm diameter. 'Bronchial' measurements were taken at the mid-point between the tracheal and peripheral airway recordings. Caecal and colonic p.d. was measured under direct vision following opening of the abdominal cavity. The composition of the bathing solution (mM) was NaCl, 140; KCl, 6; MgCl₂, 1; CaCl₂, 2; glucose, 10; HEPES, 10, pH 7.4, in all experiments, except in low-chloride solution where NaCl was replaced by equimolar sodium gluconate (measured chloride was 11 mM). The junctional potential between the normal bathing solution and the low-chloride solution was ignored because of paired comparisons. It is unlikely that the paracellular shunt is affected in the tracheal epithelium of cystic fibrosis patients²³. Forskolin (1 μM) and zardaverine (100 μM) were dissolved in ethanol and dimethyl sulphoxide, respectively, and were perfused in the presence of amiloride (100 μM). The combination of ethanol and dimethylsulphoxide has been shown not to alter murine caecal and tracheal p.d. significantly¹⁹. The measured murine nasal p.d. profile mimics that of human nasal epithelium¹⁹ and data given are obtained from the maximum measured values. Comparison of the data was made using the Mann-Whitney *U*-test and all values are expressed as the mean ± s.e.m. for convenience. Previously recorded nasal p.d.s measured using the above technique in MF1 mice showed a mean of -15.1 mV (s.e.m. ± 1.7, n=9) and rectal p.d.s of -16.7 mV (s.e.m. ± 1.3, n=7). For comparison, human nasal p.d. values are: mean of normals, -19 mV (range -2 to -36 mV, n=145); mean of cystic fibrosis patients, -46 mV (range -33 to -77 mV, n=60)³⁵. The murine tracheal p.d. profile also mimicked that in man¹⁶. Comparable values in unaffected subjects²¹ are: trachea, -16 mV (s.e.m. ± 2); bronchial airway, -11 mV (s.e.m. ± 1), and peripheral airway, -6 mV (s.e.m. ± 1), n=23. FZ, Forskalin plus zardaverine.

* *P* < 0.05; ** *P* < 0.01.

FIG. 2 Histopathology of *cf/cf* mutant mice. *a*, Homozygote *cf/cf* colon. Distension of the epithelial cells lining the crypts of colonic mucosa can be seen, caused by mucin accumulation in the cytoplasm. The glandular architecture is disrupted and the cell nuclei are compressed. *b*, Heterozygote, *cf/+* colon. The normal colonic mucosa is composed of tubular glands, lined by columnar epithelium. There is no mucin accumulation or cellular distortion. *c*, Homozygote *cf/cf* vas deferens. The lumen of the vas deferens is occupied by densely eosinophilic mucin. *d*, Heterozygote *cf/+* vas deferens. No mucin accumulation has occurred. *e*, Homozygote *cf/cf* lung. There is collapse of the lung parenchyma with mucus obstruction of a small bronchus. Lesions of this type were focal and most of the lung was morphologically normal. *f*, Heterozygote, *cf/+* lung. Morphologically normal inflated lung with patent bronchi. Images of *cf/cf* and *cf/+* mice were taken at the same magnification.

METHODS. Tissues were fixed in 10% neutral buffered formalin and then dehydrated through increasing alcohol series, cleared in xylene, and infiltrated with paraffin wax. Paraffin blocks were sectioned and slides stained with haematoxylin and eosin.



we have created an E14 line carrying the most common mutation, $\Delta F508$ (P.D., B.J.S., W.L.K., D.J.P. and J.R.D., manuscript in preparation), paving the way for further refinement of the model described here. An intrinsic property of insertional disruption that may attenuate expected phenotypes merits further comment. There is an obligate partial duplication of genomic sequences following insertion by homologous recombination, and it is possible for abnormal splicing and exon skipping to produce a normal transcript. A low level of exon skipping, generating normal transcripts, was detected after insertional disruption of the *N-myc* gene, but nevertheless the mutation generated a distinct lung defect and was perinatally lethal³¹. In contrast, no evidence for aberrant splicing was found using sensitive criteria in cultured embryonic stem cells following insertional disruption of the gene encoding hypoxanthine phosphoribosyl transferase²⁹. In our mice, the ability to distinguish heterozygotes from wild-type and mutant mice by rectal potential difference measurements argues that there can only be minimal functional protein. It will be valuable to explore *in vivo* the pharmacological possibilities for increasing chloride ion channel activity through overstimulation of any residual protein, or by alternative ion channel pathways³². We are also now in a position to modulate the level of functional CFTR by crossing these mice with other transgenic mice carrying missense mutations.

Since submission, *Cftr* mutant mice derived by a targeted gene replacement have recently been reported^{33,34}. These mice are the result of a different targeting strategy from ours and create subtly different disruptions of exon 10. Essentially the same electrophysiological method established that, like the replacement mutant³⁴, our insertional mutant was a phenotypic

'null'. The two types of mouse mutant do however display interesting and significant differences in viability and pathology which merit comment. The *Cftr* mutants created by gene replacement show frequent perinatal death with intestinal obstruction, but not lung or gonadal pathology³³. By contrast, our insertional mutants are viable and show mild intestinal obstruction in association with gut, lung and gonadal pathology, characteristic of cystic fibrosis. There are two genetic factors that may explain these phenotypic differences. First, only in the case of an insertional disruption is there the potential for production of wild-type mRNA by aberrant splicing. Under standard multiplex RNA polymerase chain reaction with appropriate internal controls, we detected reduced levels of product specific for normal message (amplification from exon 9 to 11) in *cf/+* heterozygotes compared with wild-type mice, but failed to detect more than trace levels of normal message in the *cf/cf* mice (data not shown). This is not inconsistent with a very low level of residual wild-type protein in the insertional mutant mouse, which might be sufficient to ameliorate the otherwise fatal intestinal abnormality and explain the development of other aspects of the disease more relevant to the human condition. Second, the replacement mutant mice were crossed onto three inbred backgrounds, whereas our insertional mutant mice were outcrossed onto an outbred MF1 background. The likely importance of secondary genetic factors on determining the phenotypic outcome in cystic fibrosis patients is discussed above and may also contribute to the differences between the two mutant mice. Further studies and direct comparisons will be necessary to clarify this issue.

In conclusion, the insertional disruption of the mouse *Cftr*

gene produces a phenotype that mirrors several important aspects of the human disorder. Mutant mice will be valuable in reaching a fuller understanding of the epigenetic and pathophysiological development of the disease. Future studies will

aid the development of new procedures, interventions and therapies (including somatic gene therapy) relevant to the management and treatment of cystic fibrosis. □

Received 21 August; accepted 2 September 1992.

1. Boat, T. F., Welsh, M. J. & Beaudet, A. L. in *The Metabolic Basis of Inherited Disease* 6th edn (eds Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D.) 2649–2680 (McGraw-Hill, New York, 1989).
2. Jackson, A. D. M. in *Cystic Fibrosis* (ed. Goodfellow, P.) 1–11 (Oxford Univ. Press, Oxford, 1989).
3. Kaplan, E. et al. *New Engl. J. Med.* **279**, 65–69 (1968).
4. Rommens, J. M. et al. *Science* **245**, 1059–1065 (1989).
5. Riordan, J. R. et al. *Science* **245**, 1066–1073 (1989).
6. Kerem, B. et al. *Science* **245**, 1073–1080 (1989).
7. Collins, F. S. *Science* **256**, 774–779 (1992).
8. Kristidis, P. et al. *Am. J. Hum. Genet.* **50**, 1178–1184 (1992).
9. Cutting, G. R. et al. *New Engl. J. Med.* **323**, 1685–1689 (1990).
10. Jones, C. T. et al. *Hum. molec. Gen.* **1**, 11–17 (1992).
11. Ivanschenco, T. E. et al. *Genomics* **10**, 298–299 (1991).
12. Shoshani, T. et al. *Am. J. Hum. Genet.* **50**, 222–228 (1992).
13. Anderson, M. P. et al. *Science* **253**, 202–205 (1991).
14. Diamond, G. et al. *J. biol. Chem.* **266**, 22761–22768 (1991).
15. Dorin, J. R. et al. *Transgen. Res.* **1**, 101–105 (1992).
16. Knowles, M., Gatz, J. & Boucher, R. *New Engl. J. Med.* **303**, 1489–1495 (1981).
17. Goldstein, J. L. et al. *Am. J. Physiol.* **254**, 718–724 (1988).
18. Quinton, P. M. *Nature* **347**, 226 (1990).
19. Smith, S. N., Alton, E. W. F. W. & Geddes, D. M. *Clin. Sci.* **82**, 667–672 (1992).
20. Willumsen, N. J. & Boucher, R. C. *Am. J. Physiol.* **261**, 319–331 (1991).
21. Alton, E. W. F. W. et al. *Eur. Resp. J.* **4**, 5–9 (1991).

22. Alton, E. W. F. W. et al. *Thorax* (in the press).
23. Widdicombe, J. H., Welsh, M. J. & Finkbeiner, W. E. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6167–6171 (1985).
24. Oppenheimer, E. H. & Esterly, J. R. in *Perspectives in Pediatric Pathology* Vol. 2 (eds Rosenberg, H. S. & Bolande, R. P.) 241–278 (Yearbook Medical Publishers, New York, 1975).
25. Schwachman, H., Kuiczyski, L. L. & Khaw, K. T. *Pediatrics* **36**, 689–699 (1965).
26. Wine, J. J. *Nature Genet.* **1**, 10 (1992).
27. Hasty, P., Rivera-Perez, J. & Bradley, A. *Molec. cell. Biol.* **12**, 1464–1474 (1992).
28. Koller, B. H. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10730–10734 (1991).
29. Deng, C. & Capocchi, M. R. *Molec. cell. Biol.* **12**, 3365–3371 (1992).
30. Hasty, P. et al. *Nature* **350**, 243–246 (1991).
31. Moens, C. B. et al. *Genes Dev.* **6**, 691–704 (1992).
32. Egan, M. et al. *Nature* **358**, 581–584 (1992).
33. Snouwaert, J. N. et al. *Science* **257**, 1083–1088 (1992).
34. Clarke, L. L. et al. *Science* **257**, 1125–1128 (1992).
35. Alton, E. W. F. W. et al. *Eur. Resp. J.* **3**, 922–926 (1990).

ACKNOWLEDGEMENTS. We thank J. D. Inglis, N. D. Hastie, I. J. Jackson, V. van Heyningen, D. J. H. Brock and H. J. Evans for support, advice and discussion; B. Wainwright for providing a genomic clone of mouse *Cftr*, and E. Ermslie, L. Dobbie, A. Peter and V. Ranaldi for technical assistance. P.D. is a Sir Halley Stewart Research Fellow, E.W.F.W.A. is a British Heart Foundation Intermediate Fellow. This work was supported by the Medical Research Council, UK, the Agricultural and Food Research Council, UK, and the Cystic Fibrosis Research Trust, UK.

LETTERS TO NATURE

Highly ionized gas in the nucleus of the active galaxy MCG-6-30-15

K. Nandra* & K. A. Pounds†

* Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

† Department of Physics and Astronomy, Leicester University, University Road, Leicester LE1 7RH, UK

FLUORESCENT emission from iron and the detection of a Compton reflection 'hump'^{1–3} in the X-ray spectra^{4,5} of many active galaxies are strong evidence for the presence of cold (nearly neutral) optically thick material in their nuclear regions. An apparent absorption edge at 8–9 keV, attributed to iron, in about half of a large number of Seyfert galaxies⁵ indicates the additional presence of warm (partially ionized) absorbing material, which could modify the emergent X-ray spectrum and explain flux-correlated changes in spectral shape^{6–8}. If this interpretation is correct, the warm absorber should give rise to a number of spectral features, particularly absorption edges, whose energy can be used to deduce the ionization state of the material. Here we report the detection of such a feature in the bright Seyfert galaxy MCG-6-30-15, in the form of absorption at 0.8 keV arising from the presence of highly ionized oxygen. This warm absorber will be difficult to detect at other wavelengths but, may contribute significantly to the overall X-ray opacity of the nuclear region.

The cold matter around active galactic nuclei (AGNs) may be identified with an accretion disc feeding the postulated central black hole that powers the Ginga spectra. The warm absorber will modify the X-ray emission, but because of the complexities involved in modelling the Ginga spectra, it has not been possible to exclude unambiguously the possibility that the observed absorption edge at 8–9 keV (ref. 5) could be due to cold iron, which has an absorption feature at 7.1 keV. Ginga and Exosat observations of MR2251–178 (ref. 6), NGC5548 (ref. 7) and NGC4051 (ref. 8) suggest that the presence of warm absorbing material is the most likely source of soft X-ray spectral changes, but again alternatives such as changes in the power-law index or a variable component of soft X-ray emission cannot be ruled out. If such highly ionized material is indeed present we would expect it to show other spectral imprints. In particular, the degree of ionization and column density implied by the Ginga observations would suggest a substantial oxygen K edge at

~0.8 keV. We have therefore used the spectroscopic capability of the Rosat Position Sensitive Proportional Counter (PSPC)⁹ to search for this feature in the X-ray spectrum of a bright AGN.

MCG-6-30-15 is a Seyfert I galaxy at redshift $z = 0.008$, which has now been extensively studied in the X-ray region. The source shows a strong iron $K\alpha$ line, as well as a spectral flattening above ~10 keV, consistent with the reflection spectrum common to Seyfert galaxies as a class^{4,5}. Spectral variability has also been observed, with a tight correlation of softness with flux^{8,10,11}. This behaviour is most readily explained in the context of the Compton reflection model by invoking time lags between the primary hard X-ray power law and the (flatter) reflected component^{5,8}. This was found to be an adequate description of the bulk of the spectral variability, if the timescale for variations in the reflected component was longer than a few days, and the intrinsic power-law photon index was steep ($\Gamma = 2.0$ –2.1). A residual anticorrelation of the apparent absorbing column density with flux was also observed⁵, however, which was indicative of changes in the ionization state of warm gas in the line of sight. Given the good evidence for a warm absorber in this Seyfert, it is natural to search for the spectral signature in the soft X-ray band.

MCG-6-30-15 was observed by Rosat on 1992 January 29 with the PSPC, which covers an energy range of 0.1–2.5 keV. The total exposure time was 8,500 s, with a total observed count rate

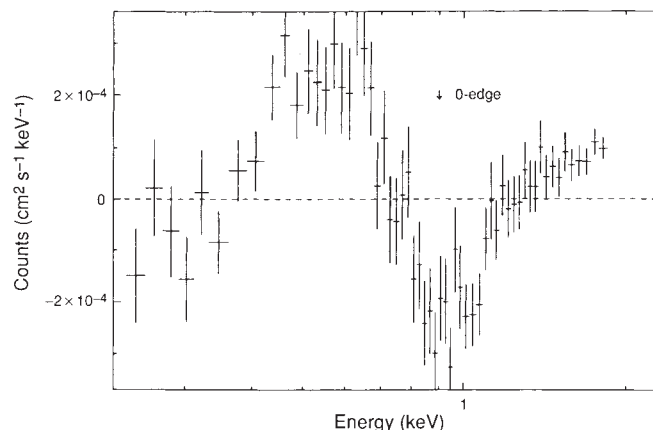


FIG. 1 Residuals from a power law fit to the Rosat PSPC soft X-ray data for MCG-6-30-15. There is a clear absorption feature at ~0.8 keV.

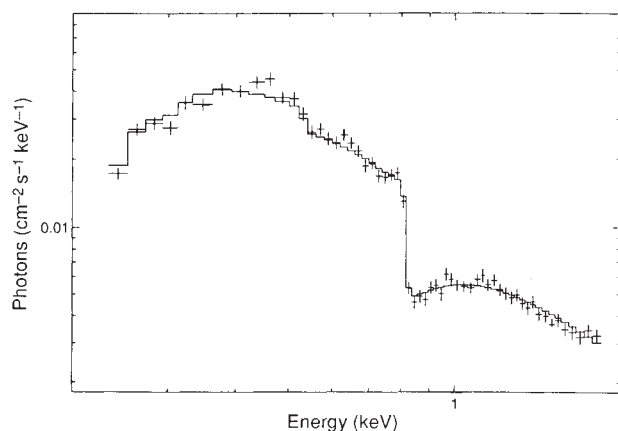


FIG. 2 Deconvolved photon spectrum derived from a fit including an absorption edge. The feature has a best-fit energy of 0.825 keV, consistent with highly ionized oxygen in warm gas in the line of sight.

of 1.8 counts s^{-1} on average, giving a high-quality spectrum. Here we consider the spectrum integrated over the whole observing period. Source photons were extracted from a circular cell with radius 3 arcmin, sufficient to include all source counts, given the 'ghost imaging' anomaly in the PSPC point spread function below ~ 0.15 keV (ref. 12). In fact, because of the relatively large Galactic column towards this source, few counts were detected below this energy anyway. The background was estimated using an adjacent source-free region of the PSPC image, 6 arcmin in radius. The subtracted spectrum, which contained $\sim 15,000$ counts in total, was corrected for dead time and mirror vignetting, divided into different bins to give at least 200 photons per channel and compared to trial models. We have used the detector response matrix released in March 1992 (S. L. Snowden, personal communication).

A simple power-law continuum, with column density, N_H , fixed at the Galactic H I value of $5 \times 10^{20} \text{ cm}^{-2}$ (ref. 13), gave an unacceptable fit to the data, with $\chi^2_\nu = 5.60$ for 57 degrees of freedom and a rather steep power-law slope of photon number index $\Gamma = 2.34$. Examination of the data-model residuals for this fit (Fig. 1) suggests that the main discrepancy arises at ~ 0.8 keV, where there is a large deficiency of counts. Adding an absorption edge to the fit reduces χ^2 by 284, significant at $\gg 99.9\%$ level on the basis of the F-statistic ($F = 238.3$; a significant improvement at the 99.9% level requires $F = 12.3$ (ref. 14)). Furthermore, the fit is now fully acceptable, with $\chi^2_\nu = 1.09$ for 55 degrees of freedom. An absorption edge is strongly preferred over an absorption notch ($\chi^2_\nu = 1.85$ at best-fit energy 0.99 keV), a gaussian absorption feature ($\chi^2_\nu = 1.76$ at 0.97 keV) or a gaussian emission feature ($\chi^2_\nu = 2.42$ at 0.38 keV). Note also that in these three cases, the spectral feature is not readily attributable to any abundant element. The deconvolved photon incident spectrum for the power law plus K-edge fit is shown in Fig. 2. The rest-frame energy of the edge is found to be 0.825 ± 0.017 keV, significantly higher than the value expected for neutral oxygen (0.533 keV), but consistent with a blend of O VII and O VIII ionization states¹⁵. The edge has a measured optical depth

$\tau = 1.20 \pm 0.15$. In addition, the power-law index derived from this fit ($\Gamma = 2.08 \pm 0.04$) is now entirely consistent with the intrinsic slope found with Ginga, supporting the interpretation of the observed hard X-ray spectrum and its variability in terms of the reflection model. The extrapolated 2–10 keV flux ($2.7 \pm 0.3 \times 10^{-11} \text{ erg s}^{-1}$) in the power law component is slightly less than that observed by Ginga in 1987 ($3.63 \pm 0.22 \times 10^{-11} \text{ erg cm}^{-2} \text{ s}^{-1}$), a relatively low state. This may well account for the oxygen absorption being so clear in the Rosat observation.

A comparison with standard photoionization models¹⁶ gives some rough values for the physical parameters of the absorbing medium. The energy of the oxygen edge, if we assume that only O VII and O VIII states are present, suggests a ratio of O VII/O VIII = 1.85, corresponding to an ionization parameter, $\xi = L/n_e R^2$, of order 100. The measured optical depth $\tau = 1.2$ implies an equivalent hydrogen column density of material $N_H \approx 2 \times 10^{22} \text{ cm}^{-2}$.

The variation in column density seen with Ginga between 1987 and 1988 (ref. 5), if attributed to changes in the ionization state of the warm medium, gives an upper limit to the recombination timescale of $t_{\text{rec}} \approx 9$ months (the separation of the two Ginga observations) and thus the gas density, $n_e \approx 3 \times 10^{11}/t_{\text{rec}}$ (ref. 17) must be greater than 10^4 cm^{-2} . The derived column density, $N_H \approx 2 \times 10^{22} \text{ cm}^{-2}$ can be equated to $n_e r$, where r is the thickness of the absorber, giving $r < 2 \times 10^{18} \text{ cm}$. The ionization parameter, ξ , gives a limit on the distance of the absorbing material. Assuming a bolometric ionizing luminosity of $L = 4 \times 10^{43} \text{ erg s}^{-1}$, appropriate for this source, we find $R < 2 \text{ pc}$ or $< 6 \times 10^6 R_g$ ($10^7 M_\odot/M_{bh}$) where $M_\odot$ is the solar mass and M_{bh} , R_g are the mass of the central black hole and the associated gravitational radius. So, although the precise location of the warm material is unclear, it seems likely that it lies close to the central continuum source. Indeed, optical observations of a similar Seyfert galaxy, Markarian 590, have shown that similar warm material may produce the very broad wings on the optical lines, and can reprocess substantial amounts of the ionizing photons into optical or ultraviolet radiation¹⁸. Alternatively, the absorption could take place in a hot scattering medium, such as that postulated in the popular Seyfert I/II unification scheme¹⁹.

Given the fact that iron K edges are indicated in $\sim 50\%$ of Seyfert galaxies observed by Ginga, it seems that substantial warm material and therefore a significant oxygen absorption edge may be common in AGN. This is supported by the tentative detection of the feature in another Seyfert I, NGC5548, based on a ~ 20 ks Rosat PSPC observation²⁰. We expect the oxygen edge to be highly variable and, together with other elements of moderate atomic mass such as neon, magnesium, silicon and sulphur, to contribute substantially to the soft X-ray opacity.

X-ray observations have now revealed both emission and absorption features in the spectra of active galaxies which indicate the presence of large amounts of gas close to the central engine. The iron K line has an origin in cold, optically thick material which may be identified with the accretion flow, whereas the oxygen edge arises in highly ionized gas. Both imply substantial reprocessing of the primary X-rays, with considerable implications for modelling of the central source and its environment. $\square$

Received 21 May; accepted 6 August 1992.

- Guilbert, P. W. & Rees, M. *Mon. Not. R. astr. Soc.* **233**, 475–484 (1988).
- Lightman, A. P. & White, T. R. *Astrophys. J.* **335**, 57–66 (1988).
- George, I. M. & Fabian, A. C. *Mon. Not. R. astr. Soc.* **249**, 352–367 (1991).
- Pounds, K. A., Nandra, K., Stewart, G. C., George, I. M. & Fabian, A. C. *Nature* **344**, 132–133 (1990).
- Nandra, K. Thesis, Leicester Univ. (1991).
- Pan, H.-C., Stewart, G. C. & Pounds, K. A. *Mon. Not. R. astr. Soc.* **242**, 177–187 (1990).
- Nandra, K. *et al. Mon. Not. R. astr. Soc.* **248**, 760–772 (1991).
- Fiore, F., Perola, G. C., Matsuoka, M., Yamauchi, M. & Piro, L. *Astr. Astrophys.* (in the press).
- Pfeffermann, E. *et al. Proc. SPIE* **733**, 519 (1986).
- Nandra, K., Pounds, K. A. & Stewart, G. C. *Mon. Not. R. astr. Soc.* **242**, 660–668 (1990).
- Matsuoka, M., Yamauchi, M., Piro, L. & Murakami, T. *Astrophys. J.* **361**, 440–458 (1990).
- Hasinger, G., Turner, T. J., George, I. M. & Boese, G. *OGIP Calibration Memo CAL/ROS/92-001*

(NASA/GSFC, Greenbelt, Maryland, 1992).

- Burstein, D. & Heiles, C. *Astr. J.* **87**, 1165–1189 (1982).
- Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences* (McGraw-Hill, New York, 1969).
- Reilman, R. F. & Manson, S. T. *Astrophys. J. Suppl. Ser.* **40**, 815–880 (1979).
- Kallman, T. R. & McCray, R. *Astrophys. J. Suppl.* **50**, 263–318 (1982).
- Osterbrock, D. E. *Astrophysics of gaseous Nebulae*, (Freeman, San Francisco, 1979).
- Ferland, G. J., Korista, K. & Peterson, B. M. *Astrophys. J.* **363**, L21–L25 (1990).
- Antonucci, R. R. J. & Miller, J. S. *Astrophys. J.* **297**, 621–632 (1985).
- Nandra, K. *et al. Mon. Not. R. astr. Soc.* (in the press).

ACKNOWLEDGEMENTS. We thank A. Fabian for his comments on the manuscript. K.N. acknowledges support of the SERC.

Evidence for sub-millisecond structure in a γ -ray burst

P. N. Bhat*, G. J. Fishman, C. A. Meegan, R. B. Wilson, M. N. Brock & W. S. Paciesas†

Space Science Laboratory, ES-62, NASA/Marshall Space Flight Center, Huntsville, Alabama 35812, USA

† Department of Physics, University of Alabama in Huntsville, Huntsville, Alabama 35899, USA

GAMMA-RAY bursts (GRBs) vary in duration from hundreds of seconds down to several milliseconds. Early studies¹ suggested that bursts with durations of <100 ms form a distinct class, accounting for a few per cent of the total number of detected bursts, and there is some evidence² for a break in the distribution of GRB durations at ~ 600 ms, perhaps implying separate physical mechanisms for long and short bursts. Recently the estimated number of short GRBs has risen substantially. The shortest burst recorded so far is GRB820405, with duration ~ 12 ms (ref. 3), and the shortest spike within a burst, an unresolved feature with width <5 ms, was in GRB841215 (refs 4–7). GRB790305 had the shortest rise-time, 0.2 ms. We report here that GRB910711, with apparently the shortest duration (~ 8 ms) yet seen by the Burst and Transient Source Experiment (BATSE), has a time profile that shows significant submillisecond structure. The responses to this burst in the different BATSE detectors, from both direct and Earth-scattered γ -rays, show that the burst is both narrower and of higher energy than is indicated by a light-curve summed over all detectors. We detected a narrow spike of duration 200 μ s in the light curve; variations on this timescale have not previously been observed in GRBs, and their explanation should be a stringent test of any GRB theory.

BATSE⁷ on the Compton Observatory has been in operation since 21 April 1991. Here we report on a burst that occurred on 11 July 1991 at 9.53 UT with features not seen previously. The location of this burst derived from relative rates from the triggered detectors is right ascension $\alpha = 14$ h 30 min and declination $\delta = -16.8^\circ$, with a location uncertainty of about 6° . Because of its low fluence, 4.7×10^{-8} erg cm^{-2} (in the energy 60–320 keV), this event was not seen by any other spacecraft, limiting location determination to the present method.

Individual time profiles of GRB910711 were derived for each of the five triggered detectors (detectors 4, 5, 6, 7 and 3) with a time resolution of 200 μ s (Fig. 1). Detector 4, with the lowest source angle, receives direct photons almost exclusively, and detector 3, with lowest Earth angle, receives almost no direct photons. The total duration of the burst as seen from the profile for detector 4 is ~ 8 ms, whereas that estimated from the light curve summed over all the detectors is ~ 16 ms. It can be seen from Fig. 1 that the detectors that record a high fraction of scattered photons (for example 7 and 3) show longer tails and their centroids are shifted in time. The relevant details are summarized in Table 1. As the position of the centroid is a shape-dependent parameter, we instead calculated the delay between the direct and delayed photons from the shift in the rising edges of the light curves of detectors 4 and 3. The delay obtained was 2.5 ± 0.3 ms and the expected delay is 2.3 ± 0.2 ms. The close agreement between the two shows that the effective scattering region is close to the Earth's surface.

The hardness ratio, defined as the ratio of the total number of signal photons above and below 100 keV, is derived for each detector. It can be seen (Table 1) that this ratio for detector 4 is nearly 4 times higher than that for detector 3 and is higher than the average value of 2.37 ± 0.14 for the signal photons summed over all the detectors.

A qualitative inspection of Fig. 1 shows several fast spikes which may be statistically significant. The most prominent is in detector 4, where 27 events fall into a single 0.2-ms bin. A time history for this detector derived with 0.1 ms resolution shows no further structure in this spike, indicating that it is fully resolved. We investigated the reality of the spike further by computing for each detector the event interval distributions during background, during the entire burst and during the spike. Counting rates were derived from a maximum-likelihood fit to the differential interval distributions and are listed in Table 1 (uncorrected for differences in detector response and recording dead time). Figure 2 shows the interval distributions for

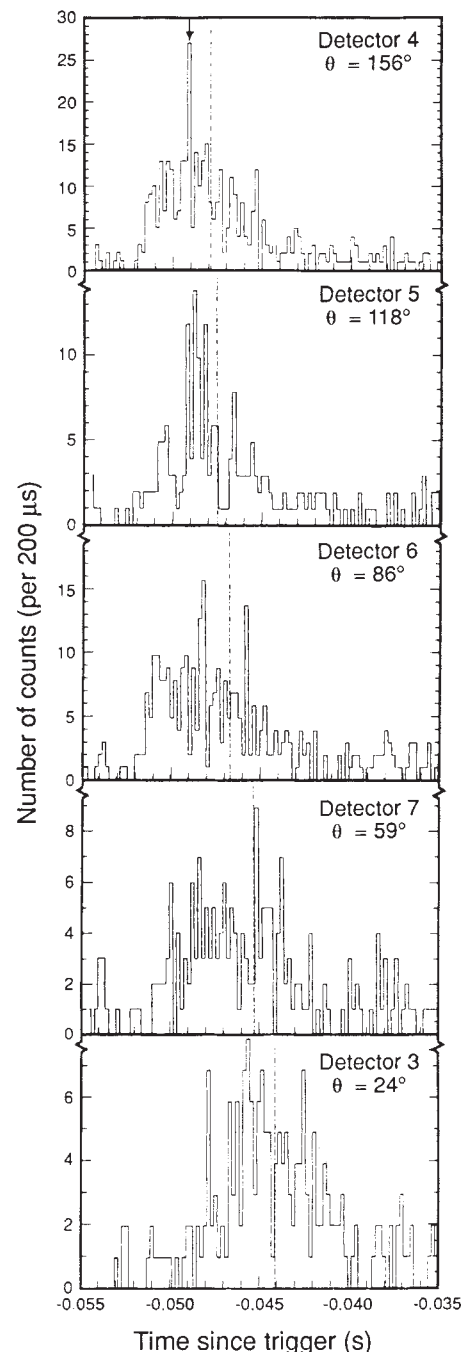


FIG. 1 Time profiles for the different triggered detectors in 200- μ s bins for energy $E > 30$ keV. The vertical dashed line shows the position of the centroid for each. The detector numbers are indicated, along with θ , the angle between the detector axis and the Earth's centre. ($\theta = 180^\circ$ corresponds to a detector pointed towards the zenith.) The spike is indicated by the arrow.

* NRC/NASA Senior Research Associate; on leave from Tata Institute of Fundamental Research, Bombay 400 005, India.

TABLE 1 Relative responses of triggered detectors to the burst GRB910711

Detector number	Source angle (°)	Earth angle (°)	Average background rate (10^3 counts s^{-1})	Average rate in burst (10^3 counts s^{-1})	Average rate in spike (10^3 counts s^{-1})	Hardness ratio	Relative delay (ms)
4	35.5	156.3	4.8	23.3	134.8	3.55 ± 0.4	0.0
5	67.7	118.0	4.2	11.7	65.4	3.51 ± 0.53	0.41
6	46.9	86.1	4.5	18.5	56.5	2.33 ± 0.28	1.00
7	75.6	58.7	4.3	10.9	16.6	1.96 ± 0.30	2.15
3	144.4	23.6	3.8	11.8	12.5	0.85 ± 0.13	3.13

detector 4. Whereas the distribution of the background photons follows the expected exponential law (Fig. 2a), the corresponding distribution for the burst is very variable (Fig. 2b). The maximum counting rate during the burst (derived from the minimum slope of the tangent to the integral interval distribution) is 4.8×10^4 counts s^{-1} . The mean number of photons from which the observed spike could have been an upwards statistical fluctuation is therefore 9.6. After the number of events has been corrected for the recording dead time of 1.875 μs per event, the probability that the spike is a statistical fluctuation is 4.5×10^{-6} , taking into account the number of trials. (This is the only burst which we have searched for short spikes.) A more conservative estimate of the chance probability can be derived by testing the null hypothesis using the 21 bins around the spike, which form the flat top of the profile, to derive the mean (spike included) which is 10.4. The probability after correcting for dead time is 7.1×10^{-5} .

We can also compare results among the detectors that viewed the burst directly. Figure 2c shows the integral interval distribution for the 27 events within the spike, which is consistent with a constant event rate of 1.35×10^5 counts s^{-1} (uncorrected for recording dead time). Table 1 lists the corresponding rates for the other detectors. The rates in detectors 4–7 during the spike are all higher than the average rate during the burst. Statistically, the results are marginally consistent with the assumption that the spike is inherent to the burst. The expected numbers (corrected for dead time) of photons in detectors 5, 6 and 7 during the spike are 17, 29 and 10, compared with the observed numbers 16, 13 and 4, respectively. We derived by χ^2 minimization an estimate of the spike intensity that is most consistent among the four detectors; this corresponds to 32 counts (deadtime corrected) in detector 4 instead of the observed 39. The minimum value of χ^2 implies consistency among the detectors only at a probability level of $\sim 6\%$.

Nevertheless, we feel that most of the evidence favours the presence of submillisecond structure in this burst. We have

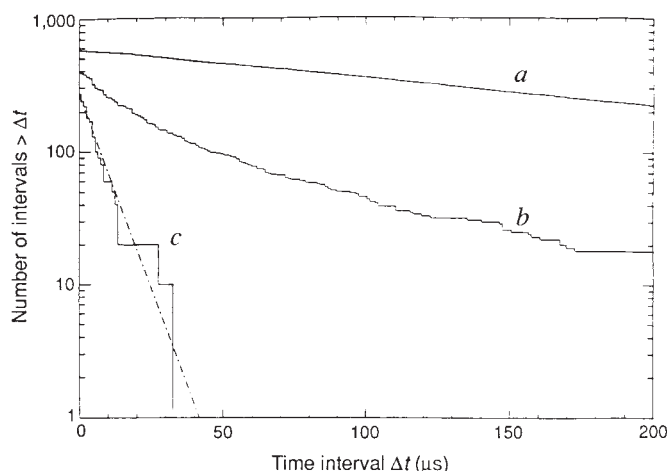


FIG. 2 Integral interval distribution for photons from detector 4 for the background (± 10 , curve a), during the burst (curve b) and the spike ($\times 10$, curve c). The variability of the count rate during the burst causes curve b to deviate significantly from a straight line. The dashed line is the maximum-likelihood fit, showing consistency with a constant count rate during the spike.

estimated the probability that the spike is a chance fluctuation by various methods, all of which yield values of at most a few times 10^{-5} , corrected for dead time and adjusted for the number of trials. The time histories show other qualitatively similar trends among the detectors, such as the low numbers of counts in bins just after the spike and high counts about five bins later. The cross-correlation function between detectors 4 and 6 shows a narrow peak at zero lag whose width is ~ 0.2 ms. Although not convincing by itself, this feature adds weight to the overall evidence. Also of relevance is the fact that we see no evidence of non-statistical fluctuations in regions of background data near the time of the burst.

The hardness ratio for the events in the spike is 6.0 ± 2.8 compared with 3.55 ± 0.4 for entire burst in detector 4. In addition, the ratios of the events in the spike to those in the rest of the burst, in energy channels 30–50 keV, 50–100 keV, 100–300 keV and > 300 keV are 0.03, 0.08, 0.08 and 0.14, respectively. The increase of these fractions with energy, together with the higher hardness ratio, indicates that the events in the spike are possibly harder.

Our results may be summarized as follows. Each triggered detector in a burst receives two types of signals: the direct signal from the source, and a signal from the photons scattered from the Earth's atmosphere which arrive after a transit-time delay. Scattering affects the temporal and spectral properties of short bursts. We have detected in GRB910711 what may be the shortest-timescale variability ever seen in a γ -ray burst. In the absence of relativistic beaming, such fast structure (~ 0.2 ms) limits the emission region to ~ 60 km, suggesting a compact source. Some other galactic models relate the observed short timescales to plasma cooling times⁸. In contrast, neutron-star models that assume that the burst photons are dominated by the upscattered thermal photons from the highly relativistic electron-positron beams predict burst rise times of a few tens of microseconds⁹. We have also searched for coherent emission of photons within the spike and find no evidence for 'photon bunching'¹⁰. Short timescales are also important constraints for cosmological models. For example, burst models involving the merger of double neutron-star binaries or neutron-star/black-hole binaries¹¹ at cosmological distances predict dynamical times, which could be related to the burst rise time, as short as 1 ms. At present, it seems that such short time structures observed in GRBs could be used only as a constraint rather than to argue for or against any model. $\square$

Received 24 March; accepted 23 July 1992.

1. Mazets, E. P. et al. *Astrophys. Space Sci.* **80**, 1–83 (1981).
2. Barat, C. et al. *Astrophys. J.* **285**, 791–800 (1984).
3. Mazets, E. P. et al. *Positron-Electron Pairs in Astrophysics* (eds Burns, M. L., Harding, A. K. & Ramaty, R.) 36–53 (Am. Inst. Phys., New York, 1983).
4. Laros, J. G. et al. *Nature* **318**, 448–449 (1985).
5. Klebesadel, R. W. et al. *IEEE Trans. Geosci. Remote Sens.* **GE-18**, 76–80 (1980).
6. Anderson, K. A. et al. *IEEE Trans. Geosci. Elect.* **GE-16**, 157–159 (1978).
7. Fishman, G. J. et al. *Proc. Gamma Ray Observatory Science Workshop*, (ed. Johnson, W. N.) 39–50 (Greenbelt, MD: NASA/GSFC, 1989).
8. Harding, A. K. *Phys. Rep.* **206**, 327–391 (1991).
9. Cherenko, A. M. & Mitrofanov, I. G. *Gamma Ray Bursts, Proc. Los Alamos Workshop on Gamma Ray Bursts, Taos, 1990* (eds Ho, C., Epstein, R. I. & Fenimore, E. E.) 55–61 (Cambridge Univ. Press, 1992).
10. Mitrofanov, I. G. *Astrophys. Space Sci.* **155**, 141–143 (1989).
11. Narayan, R. et al. *Center for Astrophysics Preprint No.* 3396 (1992).

ACKNOWLEDGEMENTS. We thank the members of the BATSE data operations team for processing the data.

Enhanced dissipation of kinetic energy beneath surface waves

Y. C. Agrawal*, E. A. Terray†, M. A. Donelan‡,
P. A. Hwang*, A. J. Williams III†, W. M. Drennan‡,
K. K. Kahma§ & S. A. Kitaigorodskii||

* Quest Integrated Inc., 21414 68th Avenue South, Kent, Washington 98032, USA

† Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

‡ National Water Research Institute, Canada Centre for Inland Waters, Burlington, Ontario L7R 4A6, Canada

§ Finnish Institute of Marine Research, SF-00931 Helsinki, Finland

|| Department of Earth and Planetary Science, The Johns Hopkins University, Baltimore, Maryland 21218, USA

TRANSFER of momentum from wind to the surface layer of lakes and oceans plays a central part in driving horizontal and vertical circulation of water masses. Much work has been devoted to understanding the role of waves in momentum transfer across the air-sea interface, but less is known about the energetics of the near-surface turbulence responsible for the mixing of momentum and mass into the underlying water column. In particular, it has remained unclear whether the structure of the turbulence in the surface layer can be described by analogy to wall-bounded shear flows or whether waves, either through breaking or wave-current interaction, introduce new length- and timescales which must be modelled explicitly. Here we report observations of turbulence in Lake Ontario, taken under conditions of strong wave breaking, which reveal a greatly enhanced dissipation rate of kinetic energy close to the air-water interface, relative to the predictions of wall-layer theory. Because wave breaking is intermittent, short-term measurements of the kinetic energy dissipation in the near-surface layer may therefore result in considerable underestimates, and any general treatment of upper mixed layer dynamics will have to take wave breaking explicitly into account.

The data presented here were acquired during 1985–1987 as part of the WAVES (Water–Air Vertical Exchange Study) programme in Lake Ontario. Observations were collected from a research tower standing in 12 m of water, 1.1 km east of the western shore of the lake. The tower was designed to minimize flow disturbances. Westerly winds produce fetch-limited waves at the tower that are of the order of 30 cm significant height. In

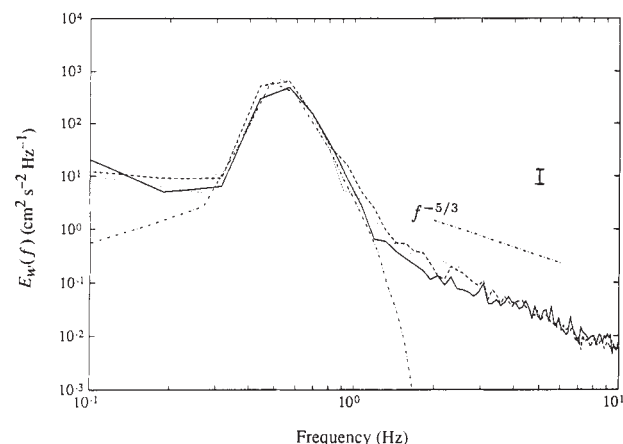


FIG. 1 Concurrent 256-s vertical velocity spectra E_w plotted against frequency from LDV at 50 cm (—), drag sphere at 50 cm (---) and BASS at 56 cm (···), with linear wave theory (50 cm) (---). The vertical band shows the 90% confidence interval. Wind speed, 11.5 m s^{-1} .

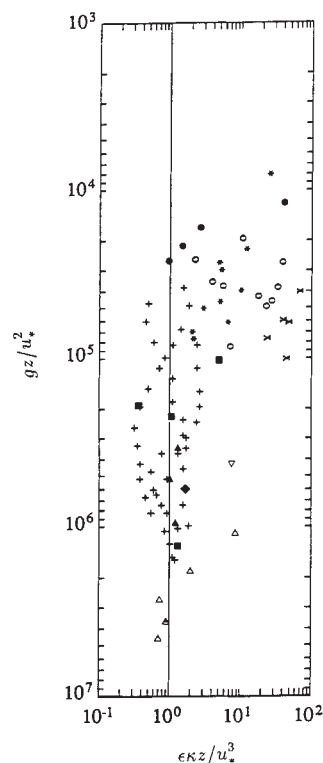


FIG. 2 Dissipation in 'wall layer' coordinates $\epsilon K z / u_*^3$ versus $g z / u_*^2$. The prediction of wall-layer theory appears as the vertical line. Data are: WAVES (drag sphere (○), BASS (*), LDV (●)), tower⁶ (■), profilers⁷ (+), tower¹¹ (△), profilers¹² (◆), tower¹³ (×), profilers¹⁴ (▲), towed hot film¹⁵ (▽), all using a common drag coefficient relation⁸.

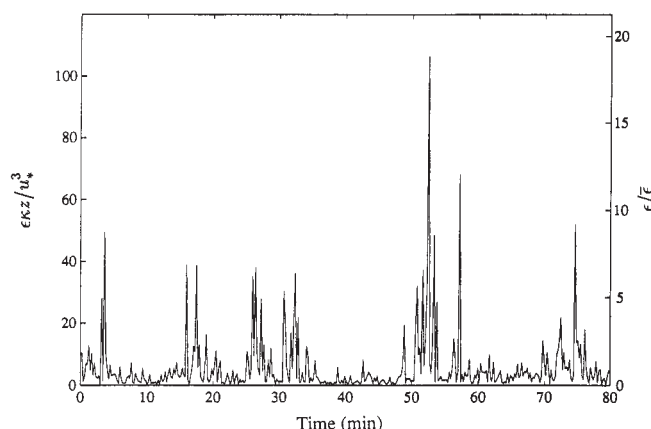
contrast, the 300 km easterly fetch can produce waves of several metres in significant height. The data presented here are from westerly, strong wind (greater than 8 m s^{-1}) cases only.

Water velocity components in the near-surface layer were measured with three different types of velocimeters mounted at different locations on the tower: an acoustic current meter array¹ (BASS), drag-sphere velocimeters² and a laser Doppler velocimeter³ (LDV). These sensors cover a range of scales: from 15 cm for BASS to 4 mm for the drag sphere and 1 mm for the LDV. BASS and drag-sphere data were continuous, lasting several hours, whereas the LDV was sequentially positioned at a range of depths between the surface and 50 cm, collecting 256-second records at each depth. In addition to the velocimeters, the tower was equipped with wave gauges, a water surface thermometer and meteorological sensors.

The dissipation results included here are derived from vertical velocity spectra, using only the high-frequency part of the spectrum well beyond the wave peak. They are obtained from the $-5/3$ slope inertial range of the velocity spectra⁴. Cross comparison of the drag-sphere, BASS and LDV spectra from roughly the same depths indicates strong consistency (Fig. 1). As these instruments are completely different in geometry and principle of measurement and were placed a few metres apart, such agreement should dispel doubts arising from possible flow disturbances contributing to the turbulence. The spectral levels at intermediate wavenumbers (between the scales of energy input and the very small scales where viscosity erodes the turbulence) reflect the energy flux from large to small scales and hence the dissipation rate. The conversion of measured frequency spectra to wavenumber space was carried out using the r.m.s. orbital velocity instead of the mean advection. In the presence of waves, it has been shown that at frequencies much higher than those of the waves, the use of r.m.s. velocity is appropriate^{4,5}.

In Fig. 2, we plot our dissipation estimates as a function of depth using data from all three instruments combined with those

FIG. 3 Time-dependent dissipation estimates, $\epsilon \kappa z / u_*^3$ (or $\epsilon / \bar{\epsilon}$), based on intervals of 13 s, the shortest data length yielding spectra with a discernible $-5/3$ slope beyond the wave peak. Wind speed: 12 m s^{-1} .



compiled in refs 6 and 7. For comparison, we present these data with the dissipation normalized using wall-layer scaling⁷. The dissipation rate ϵ is scaled with the friction velocity u_* and depth z as $\epsilon / (u_*^3 / \kappa z)$, and is unity for a wall layer ($u_* = \sqrt{\tau / \rho}$, where τ is the frictional stress at the surface due to the wind, κ is von Karman's constant and ρ is the water density; u_* is significant in that it represents a characteristic velocity of the turbulence). The depth is scaled with u_*^2 / g which is proportional to the r.m.s. height of fully developed wind-forced waves. The wind stress is derived from the measured wind based on a widely accepted formula⁸. As the state of development of the wave field was not supplied with most of the data sets taken from the literature, we omit explicitly identifying this effect, although it is likely that some of the scatter in Fig. 2 is due to the dependence of the relationship between wind and stress on wave development. The data encompass a wide range of sampling periods—from fractions of a minute to over an hour. Ninety-minute estimates of average dissipation, such as those obtained by the BASS and drag-sphere instruments, have a standard error of roughly 30%; error estimates for other cases were generally not available. Clearly our near-surface data are not well represented by wall-layer scaling (a suitable scaling will be addressed in a subsequent publication). Figure 2 presents a reasonably consistent picture of high near-surface dissipation blending into wall-layer values at greater depths.

The profiler measurements of Soloviev⁷ lie below ours for small values of gz / u_*^2 . In part this may be due to the scaling because at the low wind speeds characteristic of Soloviev's data⁷ wall-layer scaling may persist to smaller values of non-dimensional depth. In addition, if the production of turbulence is intermittent—for example, through wave breaking—then the short time samples of the profiling devices will tend to underestimate the mean dissipation rate. For illustration we explore the intermittency of the observed dissipation computed from the spectral level of the inertial sub-range in the drag sphere data observed at 1 m depth. Similar intermittency was observed by the LDV and BASS. A time history of dissipation estimates using a 13-s averaging time is shown in Fig. 3. The left-hand ordinate of the plot shows the measured dissipation rate, ϵ , normalized by the corresponding wall-layer value; the right-hand ordinate indicates the ratio of ϵ to its mean over the entire record. The probability distribution of the dissipation rate is roughly log normal. We emphasize that we are discussing the variability of 13-s averages of dissipation estimated from spectral levels and not the instantaneous dissipation that is associated with a log-normal distribution in the classical Kolmogorov theory of turbulence. It is immediately apparent that the dissipation estimates are highly intermittent with 14 values in excess of 5 times the mean, and one value of 18 times the mean. The spectral sampling error in each estimate of dissipation is $\sim 15\%$, and is much smaller than the observed intermittency. The mode of the distribution corresponds to an estimate of $1/5$ the mean,

and is approximately the expected wall-layer value. We note here that the LDV sampling time was 55 min (13 sets of 256-s records) and that of the BASS and drag spheres at least 90 min, whereas typical profilers rise rapidly through the upper layers⁷. The profilers spend at most a second or two in the wave zone and thus yield essentially random instantaneous samples from a distribution with far greater kurtosis than that of Fig. 3. Therefore an objective observer, tossing profilers through the wave zone, would, in this case, recover a value for the dissipation in general agreement (a factor of 3) with the wall-layer estimate 5 times out of 10 tosses.

The overall character of the time-dependent dissipation estimates (Fig. 3) suggests that dissipation in the upper layers is fed by energy production from two processes. The background level of dissipation is close to the wall-layer estimates and presumably arises from either the micro-breaking of small waves⁹, whose phase velocity is close to the friction velocity in the air, U_* , or, in the case of light winds, by the breakdown of the shear layer at the surface driven by the wind stress (τ) acting on the surface drift current, whose speed is also comparable to U_* (ref. 10). In either event, the energy flux from the atmosphere will be of the order of τU_* . The intense events seen in Fig. 3 result from an additional intermittent source of turbulence and, as discussed above, dominate the average. The net dissipation rates observed in strong winds at short fetch require a much higher energy flux from the wind than the wall-layer estimate of τU_* . Although we cannot identify the source of these events with certainty, one likely possibility is that they are associated with the breaking of larger waves (that is, whitecaps). This interpretation is supported by energy flux estimates based on the delivery of energy to steep waves, at frequencies above the peak of the spectrum, which yield results consistent with our observations of enhanced kinetic energy dissipation near the surface. □

Received 24 March; accepted 20 July 1992.

- Williams III, A. J. *Mar. Geol.* **66**, 345–355 (1985).
- Donelan, M. A. & Motyka, J. *Rev. scient. Instrum.* **49**, 298–304 (1978).
- Agrawal, Y. C. & Belting, C. J. *Deep Sea Res.* **35**, 1047–1067 (1988).
- Lumley, J. L. & Terray, E. A. *J. phys. Oceanogr.* **13**, 2000–2007 (1983).
- Terray, E. A. & Bliven, L. F. in *The Ocean Surface* (eds Toba, Y. & Mitsuyasu, H.) 395–400 (Reidel, Dordrecht, 1985).
- Jones, I. S. F. in *The Ocean Surface* (eds Toba, Y. & Mitsuyasu, H.) 437–442 (Reidel, Dordrecht, 1985).
- Soloviev, A. V., Vershinsky, N. V. & Bezverchniy, V. A. *Deep Sea Res.* **35**, 1859–1874 (1988).
- Large, W. G. & Pond, S. *J. phys. Oceanogr.* **11**, 324–336 (1981).
- Longuet-Higgins, M. S. *J. Fluid Mech.* **240**, 659–679 (1992).
- Wu, J. *J. Fluid Mech.* **68**, 49–70 (1975).
- Arsen'yev, S. A., Dobroklonsky, S. V., Mamedov, R. M. & Shelkovnikov, N. K. *Izv. atmos. ocean. Phys.* **11**, 530–533 (1975).
- Dillon, T. M., Richman, J. G., Hansen, C. G. & Pearson, M. D. *Nature* **290**, 390–392 (1981).
- Kitaigorodskii, S. A., Donelan, M. A., Lumley, J. L. & Terray, E. A. *J. phys. Oceanogr.* **13**, 1988–1999 (1983).
- Oakey, N. S. & Elliott, J. A. *J. phys. Oceanogr.* **12**, 171–185 (1982).
- Stewart, R. W. & Grant, H. L. *J. geophys. Res.* **67**, 3177–3180 (1962).

ACKNOWLEDGEMENTS. We acknowledge financial support from the NSF, ONR and PERD.

Short residence time of colloids in the upper ocean estimated from ^{238}U – ^{234}Th disequilibria

S. Bradley Moran & Ken O. Buesseler

Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

RECENT observations of abundant, nonliving, submicrometre particles in the upper ocean^{1–3} and new measurements of 'dissolved' organic carbon^{4–7} have fuelled speculation concerning the role of colloidal matter in ocean chemistry and biology. Colloids may act as reactive intermediates in the marine geochemistry of trace metals^{8–12}, and a biologically labile pool of colloidal matter would affect models of ocean carbon cycling^{13–16}. Here we report the use of naturally occurring ^{234}Th as an *in situ* tracer to estimate the residence time of colloidal matter in the surface waters near Bermuda. The ^{234}Th activity of colloidal matter (size range 10,000 nominal molecular weight to 0.2 μm) is similar to that of small particles (0.2–53 μm). Modelling of our results indicates a mean residence time of colloidal ^{234}Th with respect to aggregation into small particles of 10 days, which is roughly the same as for small-particle ^{234}Th , yet a factor of ~ 6 less than for the dissolved pool. These results suggest that, more generally, macromolecular colloidal matter has a short residence time and hence a rapid turnover rate in the upper open ocean.

The boundary between the dissolved and particulate phase in the oceans has traditionally been defined operationally using filters of 0.2–1 μm pore size. Owing to their small size, of order 1 nm to 1 μm (ref. 17), and flexible nature¹, colloids can pass through such filters and thereby be included in the 'dissolved' (<0.2–1 μm) phase. Quantifying the role of colloids in oceanic biogeochemical processes is particularly important if we are to understand the dynamics of the marine carbon cycle. Here we used cross-flow filtration¹⁸ to sample the colloidal size class directly. Furthermore, measurements of the ^{234}Th isotopic signature in dissolved, colloidal, and particulate phases reported here allow estimates of the timescales over which these size classes are cycled in the surface open ocean. Thorium-234 has a strong chemical reactivity for marine particle surfaces^{19–21} and a well constrained radiogenic source function²². In addition, with a half-life of 24.1 days, ^{234}Th is an excellent tracer for quantifying rates of particle cycling on timescales of ~ 1 –100 days^{19–21}, which correspond to rates of biogenic processes occurring within the surface euphotic zone.

We collected seawater samples from the upper ocean at the United States Joint Global Ocean Flux Study (JGOFS) Bermuda

Atlantic time-series (BATS) station located in the western North Atlantic (31° 50' N, 64° 10' W) during May 1991. Sea water was pumped from discrete depths onboard through 53- μm Nitex screen (1,300–3,100 l) to collect large particles, sequentially through 1- μm Nuclepore and 0.2- μm Gelman filters (240–290 l) to collect small particles, and a subsample of the filtrate (80–100 l) collected in acid-cleaned polyethylene containers. Filtration through 0.2- μm filters removes bacteria (evident from bacterial counts) and all larger microorganisms. Pre-filtered (<0.2 μm) sea water was ultrafiltered using a Millipore cross-flow filtration system modified for oceanic trace-metal sampling¹⁸ with a filter of 10,000 nominal molecular weight (NMW), equivalent to a pore size of ~ 1 –10 nm (ref. 18). The high flow rate (~ 25 l h⁻¹) of the cross-flow system facilitated the collection of large-volume (80–100 l) samples required for radiochemical determination of ^{234}Th in the colloidal and dissolved (<10,000 NMW) phase. Cross-flow filtration of each sample was completed in 4–5 hours and all samples were processed within 48 hours of collection.

Sorptive loss of ^{234}Th to the container walls and cross-flow system was <2% of the total ^{234}Th activity (determined by leaching with 1 M HCl). This is consistent with mass-balance results that indicated that total ^{234}Th activities (sum of all size classes) were within 10% of activities determined independently using large-volume sampling and γ -ray counting techniques²⁰. By comparison, a study of colloidal thorium isotopes in the shelf and slope waters in the Gulf of Mexico¹¹ reported sorptive losses of 10–40% of total ^{234}Th . Several factors may account for our relatively low sorptive losses. First, our cross-flow filter is made of regenerated cellulose which has lower sorption characteristics¹⁸ than the polysulphone filter used in the Gulf of Mexico study¹¹. Furthermore, the surface area of our Millipore filter (0.46 m²) and system walls¹⁸ is roughly half that of the Amicon system (0.88 m² filter)¹¹. In addition, loss of colloids and associated ^{234}Th due to fouling of the cross-flow filter is expected to be less problematic in open ocean waters than in particle-rich shelf and slope waters.

Vertical profiles of ^{234}Th activity (Fig. 1a) are characterized by low dissolved values and elevated colloidal plus particulate levels in the upper 100 m relative to the deeper water. The deficiency of total ^{234}Th relative to ^{238}U (Fig. 1a) in the more biologically productive upper 100 m (Fig. 1c) is consistent with scavenging and export of ^{234}Th by biologically produced colloidal and particulate matter. Colloidal ^{234}Th activities range from 0.15–0.20 d.p.m. l⁻¹ and show no systematic variation over the depth range sampled (Fig. 1b). ^{234}Th activities in the 0.2–1 μm particulate fraction are high in the near-surface waters and decrease toward the base of the euphotic zone. A broad maximum in ^{234}Th activity is evident in the 1–53 μm size class, slightly below the chlorophyll *a* maximum. Large-particle

FIG. 1 Vertical profiles of a, ^{234}Th activity on the sum of the colloidal and particulate phases (●, >10,000 NMW), the dissolved phase (○, <10,000 NMW) and the total (■, sum of all size classes) and dissolved ^{238}U activity²² (dotted line); b, ^{234}Th activity on colloidal (●, 10,000 NMW to 0.2 μm), small particle (○, 0.2–1 μm ; △, 1–53 μm), large particle (▲, >53 μm), and the sum of the >10,000 NMW (■, colloidal plus small and large particle) size classes; c, chlorophyll *a* (●) and particulate organic carbon (○) in the upper ocean at the BATS station (31° 50' N, 64° 10' W), 14 May 1991. Colloidal and particulate ^{234}Th activities were determined by ion-exchange and beta-counting procedures³³; dissolved ^{234}Th was collected on MnO_2 impregnated cartridges and quantified using γ -ray counting techniques²⁰. Error bars (sometimes smaller than the data symbol) were determined from counting statistics. Particulate organic carbon and chlorophyll *a* concentrations were determined using standard methods³⁴.

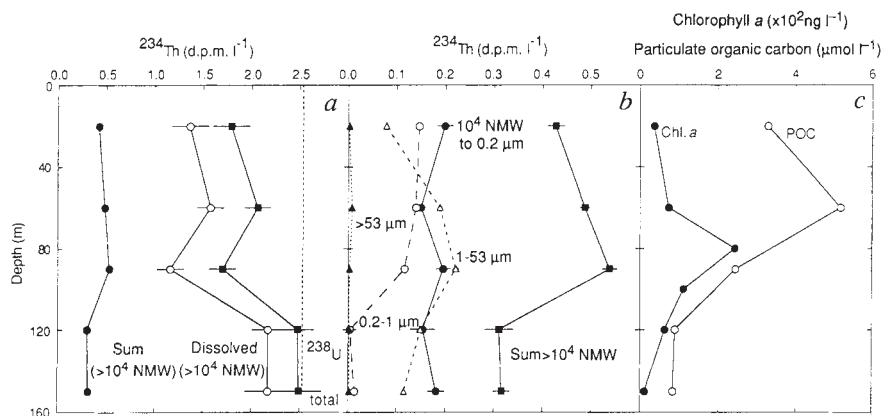


TABLE 1 Average of size-fractionated ^{234}Th results from 0–150 m off Bermuda

Quantity	Dissolved	Colloidal	Small particle	Large particle
^{234}Th (d.p.m. l^{-1})	1.62 ± 0.10	0.18 ± 0.01	0.25 ± 0.01	0.0040 ± 0.0002
τ (days)	61.3 ± 7.4	9.6 ± 1.4	18.3 ± 3.8	0.3 ± 0.1
Percentage of particulate	—	41.5 ± 2.7	57.6 ± 3.0	0.9 ± 0.1
Percentage of total	78.8 ± 6.2	8.8 ± 0.7	12.2 ± 0.8	0.2 ± 0.1

Size classes defined as: dissolved, <10,000 NMW; colloidal, 10,000 NMW to $0.2\text{ }\mu\text{m}$; small particle, $0.2\text{--}53\text{ }\mu\text{m}$; large particle, $>53\text{ }\mu\text{m}$. Errors were propagated from counting statistics and average ^{234}Th inventories (0–150 m) using trapezoidal integration. Results from 14 May 1991.

($>53\text{ }\mu\text{m}$) ^{234}Th activities are extremely low ($0.002\text{--}0.008\text{ d.p.m. l}^{-1}$) and correlate to some extent with particulate organic carbon ($r^2 = 0.67$, $n = 5$).

A striking aspect of the size-fractionated data is that, on average, ^{234}Th activities in the colloidal size range and on small particles are similar in magnitude (Table 1). As a percentage of the total ^{234}Th activity, colloids and small particles each constitute $\sim 10\%$, with the balance primarily in the dissolved (<10,000 NMW) phase. The implication of these results is that oceanic colloidal matter may be as important as traditionally defined suspended particulate matter ($0.2\text{--}53\text{ }\mu\text{m}$) in the cycling of Th^{19–24}, and possibly other reactive elements^{9,12,25–27}, between dissolved and particulate forms.

We calculated mean residence times of ^{234}Th in the dissolved, colloidal and particulate phases using an existing scavenging model^{19,20} that includes colloidal and large particle size classes (Fig. 2). The principal model assumptions are that the ^{234}Th activity is at steady state, scavenging of ^{234}Th is essentially irreversible, and advection and diffusion terms are negligible relative to net removal fluxes^{19,20}. The activity of dissolved ^{234}Th is controlled by a balance between production from its soluble parent, ^{238}U (half-life 4.5×10^9 years), loss by radioactive decay, and removal on colloids and larger particles. Colloid–particle interactions are simplified as a sequential process such that the ^{234}Th activity in each pool is maintained by production from a smaller size class, loss by radioactive decay and a net transfer to the next larger size class. Rapid export of ^{234}Th from the surface waters occurs only on large, fast-sinking aggregates. With these assumptions, the differential equations describing scavenging of ^{234}Th in the upper ocean are

$$\partial A_{\text{Th}}^{\text{d}} / \partial t = 0 = A_{\text{U}} \lambda - A_{\text{Th}}^{\text{d}} \lambda - J_{\text{Th}}$$

$$\partial A_{\text{Th}}^{\text{c}} / \partial t = 0 = J_{\text{Th}} - A_{\text{Th}}^{\text{c}} \lambda - C_{\text{Th}}$$

$$\partial A_{\text{Th}}^{\text{p}} / \partial t = 0 = C_{\text{Th}} - A_{\text{Th}}^{\text{p}} \lambda - P_{\text{Th}}$$

$$\partial A_{\text{Th}}^{\text{l}} / \partial t = 0 = P_{\text{Th}} - A_{\text{Th}}^{\text{l}} \lambda - L_{\text{Th}}$$

where A_{U} is the activity of ^{238}U ($0.069 \times \text{salinity}^{22}$), λ is the decay constant for ^{234}Th (0.0288 day^{-1}) and A_{Th} is the activity (d.p.m. l^{-1}) of ^{234}Th in dissolved (d), colloidal (c), small-particle (p), and large-particle (l) size classes. The terms J_{Th} , C_{Th} , P_{Th}

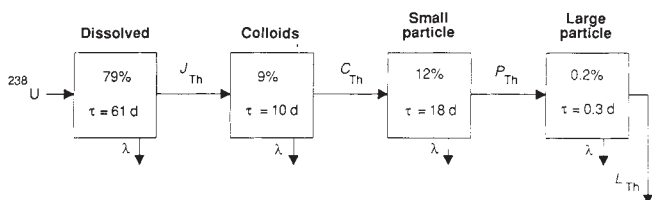


FIG. 2 Particle cycling model used to calculate mean residence times of ^{234}Th in dissolved, colloidal, small particle, and large particle size classes in the upper ocean. J_{Th} , C_{Th} , P_{Th} and L_{Th} represent net removal fluxes of ^{234}Th from the various pools and λ represents the radioactive decay constant for ^{234}Th . Also shown is the percentage of the total ^{234}Th activity associated with each size class and the corresponding residence time.

and L_{Th} represent net removal fluxes (d.p.m. $\text{l}^{-1}\text{ day}^{-1}$) of dissolved, colloidal, small- and large-particle ^{234}Th , respectively.

Clearly, interactions between colloids, small particles, and large particles are complex²⁸ and presumably mediated by a combination of biotic (bacterial and other microorganism activity, zooplankton grazing) and abiotic (adsorption, aggregation, disaggregation) processes. The critical aspect of our model that determines colloidal ^{234}Th residence times is the coagulation of colloids with small particles. It is conceivable that additional interactions, for example scavenging of dissolved ^{234}Th directly onto the surfaces of picoplankton and fine detritus, may operate to bypass the colloidal pool; such interactive effects could be investigated using additional Th isotopic information^{23,24}. Recent theoretical⁸ and laboratory¹⁰ studies, however, support our model assumptions that coagulation is the rate-limiting step for the transfer of reactive metals from the dissolved to the particulate phase. Moreover, the simplified scavenging model used here allows the calculation of residence times using a single Th isotope^{19–21}.

The mean residence time (τ) of ^{234}Th in the i th size class is calculated using $\tau_i = A_{\text{Th}}^i / R_i$, where A_{Th}^i is the ^{234}Th activity and R_i is the net removal flux. We calculated residence times for each size class using average ^{234}Th values for the upper 150 m (summarized in Table 1). The mean residence time of ~ 10 days calculated for ^{234}Th in the colloidal fraction is within a factor of 2 of the residence time of ^{234}Th on small particles and 6 times less than for the dissolved pool. The extremely short residence time of ^{234}Th on large particles (<1 day) is consistent with their rapid export from the surface waters.

One of the intriguing aspects of the recent 'dissolved' organic carbon (DOC) data is the large vertical and horizontal gradients in the surface oceans^{4,5}. Although the amount of DOC in the oceans is at present controversial^{4–7}, such spatial variability is unlikely to be eliminated because of DOC blank corrections²⁹. The observed gradients imply that DOC is produced and remineralized more rapidly than previous estimates suggested^{13–16}. Our size-fractionated ^{234}Th results allow for an estimate of the time-scale over which high-molecular-weight colloidal matter is cycled in the upper ocean. In relating the residence time of colloidal ^{234}Th to colloids in general, we must assume that ^{234}Th adsorbs on all colloids. Given the high reactivity of ^{234}Th for oceanic particle surfaces^{19–21}, this is a reasonable approximation. The implication of our results is that, on average, the high-molecular-weight component of 'dissolved' organic carbon is cycled on the order of $\sim 1\text{--}2$ weeks in the surface open ocean. Moreover, such high colloidal turnover rates have recently been reported in the Gulf of Mexico¹¹.

The mean residence time of the colloidal size class reported here is comparable to DOC residence times of 30–40 days estimated from incubation experiments and of 10–160 days determined *in situ* during a spring phytoplankton bloom³⁰. Such rapid turnover times are considerably shorter than the previous estimate for the mean apparent ^{14}C age of DOC in the upper ocean of $>1,000$ years³¹, further suggesting that some fraction of DOC is actively involved in carbon cycling in ocean surface waters³¹. Indeed, Benner *et al.*⁶ report that high-molecular-weight components of dissolved organic matter are reactive and support much of the heterotrophic activity in the surface ocean.

Thus, the size-fractionated ^{234}Th results reported here provide new evidence for a more dynamic role of 'dissolved' organic carbon in the ocean carbon cycle than previously believed³². □

Received 30 March; accepted 31 July 1992.

- Koike, I., Hara, S., Terauchi, K. & Kogure, K. *Nature* **345**, 242–244 (1990).
- Wells, M. L. & Goldberg, E. D. *Nature* **353**, 342–344 (1991).
- Longhurst, A. R. *et al. Deep-Sea Res.* **39**, 1–7 (1992).
- Sugimura, Y. & Suzuki, Y. *Mar. Chem.* **24**, 105–131 (1988).
- Martin, J. H. & Fitzwater, S. E. *Nature* **356**, 699–700 (1992).
- Benner, R. *et al. Science* **255**, 1561–1564 (1992).
- Ogawa, H. & Ogura, N. *Nature* **356**, 696–698 (1992).
- Honeyman, B. D. & Santschi, P. H. *J. mar. Res.* **47**, 951–992 (1989).
- Moran, S. B. & Moore, R. M. *Geochim. cosmochim. Acta* **53**, 2519–2527 (1989).
- Honeyman, B. D. & Santschi, P. H. *Envir. Sci. Tech.* **25**, 1739–1747 (1991).
- Baskaran, M., Santschi, P. H., Benoit, G. & Honeyman, B. D. *Geochim. cosmochim. Acta* (in the press).
- Wells, M. L. *et al. Nature* **353**, 248–250 (1991).
- Toggweiler, J. R. *Nature* **345**, 203–204 (1990).
- Toggweiler, J. R. *Productivity of the Oceans: Present and Past* (eds Berger, W. H. *et al.*) 65–84 (Wiley, New York, 1989).
- Jackson, G. W. *Oceanography* **1**, 28–33 (1988).
- Najjar, R. G., Sarmiento, J. L. & Toggweiler, J. R. *Global biogeochem. Cycles* **6**, 45–76 (1992).

- Stumm, W. *Envir. Sci. Tech.* **11**, 1066–1070 (1977).
- Moran, S. B. *Marine Particles: Analysis and Characterization* (eds Hurd, D. C. & Spencer, D. W.) 275–280 (AGU, Washington, 1991).
- Coale, K. H. & Bruland, K. W. *Limnol. Oceanogr.* **32**, 189–200 (1987).
- Buesseler, K. O. *et al. Deep-Sea Res.* **39**, 1115–1137 (1992).
- Buesseler, K. O. *Nature* **353**, 420–422 (1991).
- Chen, J. H., Edwards, R. L. & Wasserburg, G. J. *Earth planet. Sci. Lett.* **80**, 241–251 (1986).
- Nozaki, Y., Horibe, Y. & Tsubota, H. *Earth planet. Sci. Lett.* **54**, 203–216 (1981).
- Bacon, M. P. & Anderson, R. F. *J. geophys. Res.* **87**, 2045–2056 (1982).
- Moran, S. B. & Moore, R. M. *Nature* **335**, 706–708 (1988).
- Moore, W. S. & Dymond, J. *Nature* **331**, 339–341 (1988).
- Fisher, N. S., Cochran, J. K., Krishnaswami, S. & Livingston, H. D. *Nature* **335**, 622–625 (1988).
- Bruland, K. W. *et al. Productivity of the Oceans: Present and Past* (eds Berger, W. H. *et al.*) 193–215 (Wiley, New York, 1989).
- Toggweiler, J. R. *Nature* **356**, 665–666 (1992).
- Kirchman, D. L. *et al. Nature* **352**, 612–614 (1991).
- Williams, P. M. & Druffel, E. R. M. *Nature* **330**, 246–248 (1987).
- Druffel, E. R. M. & Williams, P. M. *Nature* **347**, 172–174 (1990).
- Fleer, A. P. & Bacon, M. P. *Marine Particles: Analysis and Characterization* (eds Hurd, D. C. & Spencer, D. W.) 227–228 (AGU, Washington, 1991).
- US Joint Global Ocean Flux Study *BATS Data Rep. B-1A* (1991).

ACKNOWLEDGEMENTS. We thank A. F. Michaels and the Bermuda Biological Station for Research, Inc. for providing the sampling opportunity and ancillary data, E. R. M. Druffel, M. P. Bacon, P. H. Santschi, M. Baskaran and B. D. Honeyman for comments, and R. M. Moore, J. E. A. Andrews and M. C. Hartman for technical assistance. Financial support was provided by NSERC of Canada, WHOI, the NOAA Sea Grant Program and the NSF.

Evidence from nitrogen isotope ratios for enhanced productivity during formation of eastern Mediterranean sapropels

S. E. Calvert*, B. Nielsen* & M. R. Fontugne†

* Department of Oceanography, University of British Columbia, Vancouver, British Columbia, Canada V6T 1W5

† Centre des Faibles Radioactivités, Laboratoire Mixte CNRS-CEA, Avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France

THE formation of horizons of organic-rich sediment (sapropels) in the eastern Mediterranean during the Holocene and Upper Pleistocene¹ has been ascribed² to enhanced preservation of organic carbon in bottom waters rendered anoxic by the restriction of deep-water renewal, either by lowered sea level during glacial maximum conditions³ or by the presence of low-salinity surface waters derived from increased runoff^{2,4}. A second possibility is that the sapropels formed as a consequence of higher settling fluxes of organic matter caused by increased primary production connected with the increased runoff^{5,6}. Here we report that in a sediment core from close to the mouth of the Nile, the sapropels have significantly lighter nitrogen isotope ratios ($^{15}\text{N}/^{14}\text{N}$) than the intercalated marl oozes. We adduce evidence that these large differences cannot be caused either by variable mixtures of marine and terrestrial organic matter having different isotopic compositions, or by differences in the extent and type of post-depositional alteration. The differences are consistent, however, with a higher utilization of dissolved nitrate during sapropel formation, implying that the sapropels were ultimately formed as a consequence of high plankton production, which led to high fluxes of organic matter to the sea floor.

Core MD 84641 (33° 02' N, 32° 38' E; 1,375 m water depth) consists of 11.61 m of pale cream to yellowish brown nanofossil and foraminiferal marl ooze with nine olive-green sapropels varying in thickness from 6 to 22 cm. The peak organic carbon contents of the sapropels range from 2 to 4.5% by weight (Fig. 1a). The marl oozes have organic carbon contents ranging from 0.1 to 0.3%, with no systematic trends in this lithology down the core. The foraminiferal isotope record of the core⁷ shows that it appears to extend back to stage 12, and that it contains sapropels S1 and S3–S10 (ref. 8).

The $\delta^{15}\text{N}$ record (Fig. 1b) shows large and systematic variations, with a total amplitude of more than 9‰. The systematically lighter values in all sapropel horizons reach 0.3‰ in S5, whereas the heaviest values are confined to the glacial stages. In detail, the variations in $\delta^{15}\text{N}$ closely follow those of organic carbon; in sapropels S3, S4, S5 and S8 there are minima in organic carbon between two maxima, and the $\delta^{15}\text{N}$ values change abruptly to heavier values at these same horizons (Fig. 2).

The detailed composition and isotope records in core MD 84641 imply that either the source or the diagenetic history of the organic matter in the core has changed between the organic-rich and organic-lean horizons over the past 450 kyr in this part of the eastern Mediterranean. Variable sources include marine planktonic and terrestrial organic matter, as well as variations in the composition of marine organic matter itself. Diagenetic effects include differences in the degree and/or rate of release of the two isotopes during organic matter degradation during settling through the water column and during burial on the sea floor.

Terrestrial contributions to the sapropels are minor because their $\delta^{13}\text{C}_{\text{organic}}$ values (mean $-21.0 \pm 0.82\text{‰}$) are identical to those of modern Mediterranean plankton and there is no gradient in the isotope values in cores at variable distances from the Nile river, the principal source of terrestrial materials⁷. Moreover, the lipids in several of the sapropels have a predominantly marine source^{9,10}. It is therefore unlikely that the isotope record shown in Fig. 1b is affected significantly by different sources of organic matter having different isotopic compositions¹¹.

The isotopic composition of sedimentary organic matter could be different from that produced by photosynthesis because of its alteration during settling through the water column and as a result of burial diagenesis. Such changes have been observed in laboratory cultures^{12–14}, implying that microbial alteration of the deposited organic matter does lead to changes in the $\delta^{15}\text{N}$ of residual organic matter. The mechanism may involve loss of isotopically distinct fractions of this material^{14–16} or kinetic isotopic fractionation, possibly by peptide bond breakage¹⁷. The isotopic composition of settling particulate matter in the ocean is known to change significantly, resulting in either an increase^{18–20} or a decrease^{19–22} in $\delta^{15}\text{N}$ with water depth. The increases observed are consistent with the preferential removal of ^{14}N in the products of microbial reactions and the concomitant enrichment of ^{15}N in the source material, whereas the decreases observed have been explained by the removal of a

nitrogen fraction enriched in ^{15}N or the gain of component(s) depleted in ^{15}N with depth²².

There have been reports of marked increases in the $\delta^{15}\text{N}$ values with burial in some near-shore sediments²³, which have been interpreted as diagenetic changes with the preferential release of the lighter isotope. Other studies, however, have shown little change in $\delta^{15}\text{N}$ with sediment depth^{24,25}, implying no significant diagenetic alteration of the isotopic composition. If the relative abundances of the two isotopes were progressively changed by diagenesis, this should appear as secular changes in the isotopic composition of the nitrogen in sediment sequences. The lack of such a trend and the presence of abrupt changes

between the two lithologies strongly suggest that diagenetic changes alone have not produced the isotope record shown in Fig. 1b.

If we can eliminate variable organic matter sources and diagenetic changes to the isotopic composition of the sediments in core MD 84641, we are left with variations in the isotopic composition of the planktonic organic matter as the main cause of the observed isotopic contrast between the two lithologies. Assimilation of dissolved nitrate and ammonia by plants results in significant isotope fractionation, the particulate product being depleted and the residual dissolved nutrient correspondingly enriched in ^{15}N (refs 13, 26). Fractionation factors for nitrate uptake by phytoplankton in culture vary from 1.000 to 1.063 under nutrient-replete conditions^{26,27}, and that for ammonia assimilation is reported to range from 1.00 to 1.01 (ref. 26). From field studies in which the isotopic composition of particulate organic nitrogen was compared with that of the nutrient source, the fractionation falls in the range 1.006–1.009 (refs 13, 22, 25, 28–30).

On the basis of this information, we suggest that the large changes in $\delta^{15}\text{N}$ in core MD 84641 probably reflect different isotope fractionations during nitrate uptake under nutrient-replete and nutrient-depleted conditions during the accumulation of the sapropels and the marl oozes, respectively. That is,

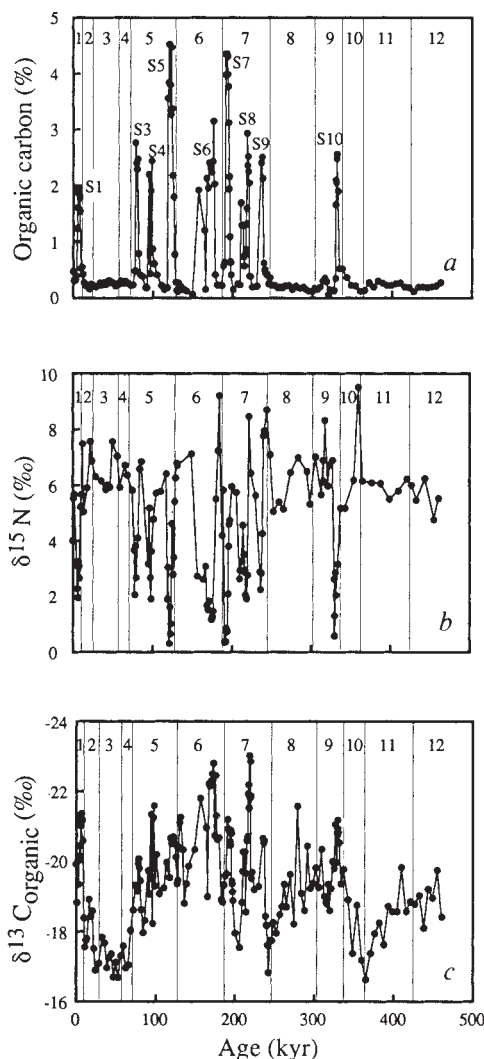


FIG. 1 *a*, Distribution of organic carbon in core MD 84641. The age model was derived from the foraminiferal stratigraphy, radiocarbon dating and magnetic reversal stratigraphy⁷. Stages are indicated at the top. The sapropels are labelled S1 and S3–S10 (ref. 50). *b*, The $\delta^{15}\text{N}$ record in core MD 84641. The samples were prepared for isotopic analysis following the methods described in ref. 23 and the determinations were made with a VG PRISM mass spectrometer. The results are reported in the δ notation, $\delta^{15}\text{N} = [(^{15}\text{N}/^{14}\text{N})_{\text{sample}} / (^{15}\text{N}/^{14}\text{N})_{\text{standard}} - 1]$, relative to atmospheric nitrogen (NBS-14) and the precision of the measurements is better than $\pm 0.2\%$. We have assumed that the isotopic composition of the total nitrogen obtained in the combustion reflects almost entirely the composition of organic nitrogen⁵¹, and that inorganic forms of nitrogen constitute a minor proportion of the total. In any case, the isotopic composition of interstitial ammonia produced from the breakdown of sedimentary organic matter seems to be very similar to the source organic matter⁵². *c*, The $\delta^{13}\text{C}_{\text{organic}}$ record in core MD 84641 from ref. 7. $\delta^{13}\text{C} = [(^{13}\text{C}/^{12}\text{C})_{\text{sample}} / (^{13}\text{C}/^{12}\text{C})_{\text{standard}} - 1]$ where standard is PDB.

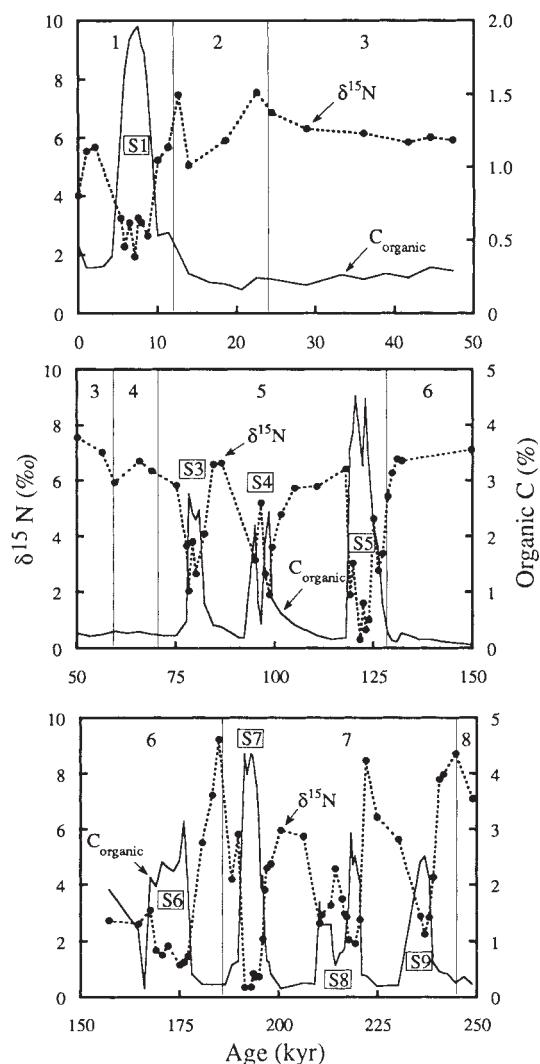


FIG. 2 Detailed $\delta^{15}\text{N}$ and organic carbon records through sapropels S1 and S3–S9 to show the sharp changes at the boundaries between the sapropels and the marl oozes. Isotope stages are shown at the top of each panel.

nitrate and/or ammonia were present at high concentration in the near-surface waters during the formation of the sapropels, and maximal isotope fractionation led to the production of relatively ^{15}N -depleted organic nitrogen which settled to the sea floor. During the formation of the marl oozes, on the other hand, the prevailing oligotrophic conditions, like those of today, resulted in the production of planktonic organic matter whose isotopic composition was heavier and closer to that of the nutrient substrate. This effect has been invoked³¹ to explain the differences in $\delta^{15}\text{N}$ of particulate organic matter between eutrophic and oligotrophic areas of the ocean; $\delta^{15}\text{N}$ is lowest where nitrate concentrations are high, whereas $\delta^{15}\text{N}$ is highest where dissolved nitrate concentrations are low. Modern surface sediment $\delta^{15}\text{N}$ reflects this control in other areas of the ocean^{25,32}. The $\delta^{15}\text{N}$ is low beneath areas of high nitrate uptake in the upwelling region along the Equator in the Pacific and in the area south of the polar front in the Southern Ocean, and increases towards regions of lower nitrate uptake (oligotrophic conditions) in both areas.

The suggestion that the $\delta^{15}\text{N}$ record in core MD 84641 reflects variable nutrient uptake implies that the nutrient content of the near-surface waters was high during the formation of the sapropels. Support for this idea is supplied by the fivefold higher Cd/Ca ratios in *Globigerinoides ruber* in the horizon below and within the youngest sapropel compared with the last glacial maximum and the recent sediments³³, suggesting that phosphate concentrations in the near-surface waters were significantly higher than today. Abundances of planktonic foraminifera that require seasonal deep mixing of the upper waters to complete their life cycles, especially *Globorotalia truncatulinoides*, are abundant just below and above some sapropels, implying that there was weak vertical stability in the eastern basin at these times³⁴. Periods of deep mixing may therefore have supplied the surface waters with deep nutrients during transitions between the present anti-estuarine circulation regime and an estuarine condition^{5,35}, the latter promoted by flooding from the Nile river^{4,36} and possibly other sources³⁷.

This interpretation would be complicated if the abundance of diazotrophic phytoplankton were significantly higher in the sapropels than in the associated marl oozes. Nitrogen fixation produces cell $\delta^{15}\text{N}$ values close to that of the atmospheric source ($\delta^{15}\text{N}=0$) because the fractionation effect for this reaction is

small³⁸⁻⁴⁰. Low $\delta^{15}\text{N}$ values in plankton samples from some nutrient-poor areas of the ocean have been ascribed to N_2 -fixing cyanobacteria^{18,19,31}, although it has also been suggested that such low values could be due to the assimilation of isotopically light regenerated ammonium under oligotrophic conditions⁴¹. Some marginal areas of the ocean affected by high freshwater runoff do have high abundances of diazotrophic phytoplankton⁴². A possible modern analogue for the eastern Mediterranean during the formation of the sapropels is in the Andaman Sea, where the freshwater input from the Irrawaddy and Salouen rivers leads to a stratified water column with a low-salinity surface layer. Here, cyanophyceae are the most abundant phytoplankton⁴³. They seem to flourish only in nutrient-depleted waters⁴⁴, however, and their presence during the sapropel events would therefore be inconsistent with the high nutrient conditions in the surface waters implied by the high foraminiferal Cd/Ca ratios in S1 (ref. 33). Moreover, the oligotrophic conditions that prevail in the eastern Mediterranean today and probably during the last glacial maximum would be more conducive to the development of a diazotrophic phytoplankton population, and if such organisms dominated the primary production one would expect the organic nitrogen to be isotopically lighter than that produced by phytoplankton that use dissolved inorganic nitrogen. But the record in core MD 84641 (Fig. 1b) is inconsistent with this effect. We therefore discount the influence of nitrogen fixation on primary production in the eastern Mediterranean during the formation of the sapropels.

The conclusion that the sapropels formed during periods of enhanced nutrient utilization has important implications for hypothesis for their origin. High nutrient utilization implies high nutrient supply to the mixed layer, which in turn implies high primary production, as has been suggested on other grounds^{45,46}. The resulting high production would have resulted in a high settling flux of organic matter, a prerequisite for the formation of the organic-rich facies⁴⁷. The high carbon flux would also have caused a decrease in oxygen in the deep waters. This is consistent with model results⁴⁸ showing that low deep-water oxygen levels in the Mediterranean can only be produced by higher deep-water nutrient concentrations. A possible mechanism for this change is a reversal in the flow directions of the surface and deep waters^{5,6,35,49} brought about by periods of high freshwater discharge. □

Received 24 April; accepted 28 July 1992.

1. Kullenberg, B. K. *Vet. Q. Vitterh. Samh. Handl.* **B6**, 3-37 (1952).
2. Olsson, E. *Rep. Swedish Deep-Sea Exped.* **8**, 337-391 (1961).
3. Bradley, W. H. *Science* **88**, 376-379 (1938).
4. Rossignol-Strick, M., Nesteroff, W., Olive, P. & Vergnaud-Grazzini, C. *Nature* **295**, 105-110 (1982).
5. Berger, W. H. in *Chemical Oceanography* (eds Riley, J. P. & Chester, R.) 265-388 (Academic, London, 1976).
6. Calvert, S. E. *Oceanol. Acta* **8**, 255-267 (1983).
7. Fontugne, M. R. & Calvert, S. E. *Paleoceanography* **7**, 1-20 (1992).
8. Cita, M. B. *et al. Quat. Res.* **8**, 205-235 (1977).
9. Smith, D. J., Eglinton, G. & Morris, R. J. *Phil. Trans. R. Soc. A* **319**, 375-415 (1986).
10. ten Haven, H. L., Baas, M., de Leeuw, J. W. & Schenck, P. A. *Mar. Geol.* **75**, 137-156 (1987).
11. Sweeney, R. E., Liu, K. K. & Kaplan, I. R. in *Stable Isotopes in the Earth Sciences* (ed Robinson, B. W.) 9-26 (DSIR Bulletin, 1978).
12. Miyake, Y. & Wada, E. *Rec. Oceanogr. Works Japan* **11**, 1-6 (1971).
13. Wada, E. in *Isotope Marine Chemistry* (eds Goldberg, E. D., Horibe, Y. & Saruhashi, K.) 375-398 (Uchida Rokuhaku, Tokyo, 1980).
14. Macko, S. A. & Estep, M. L. *F. Org. Geochem.* **6**, 787-790 (1984).
15. Macko, S. A., Estep, M. L. F., Engel, M. H. & Hare, P. E. *Geochim. cosmochim. Acta* **50**, 2143-2146 (1986).
16. Macko, S. A., Estep, M. L. F., Hare, P. E. & Hoering, T. C. *Chem. Geol.* **65**, 79-92 (1987).
17. Silfer, J. A., Engel, M. H. & Macko, S. A. *Isotope Geosci.* (in the press).
18. Saino, T. & Hattori, A. *Nature* **283**, 752-754 (1980).
19. Saino, T. & Hattori, A. *Deep-Sea Res.* **34**, 807-827 (1987).
20. Altabet, M. A. *Deep-Sea Res.* **35**, 535-554 (1988).
21. Wada, E., Terazaki, M., Kabaya, Y. & Nemoto, T. *Deep-Sea Res.* **34**, 829-841 (1987).
22. Altabet, M. A., Deuser, W. G., Honjo, S. & Stein, C. *Nature* **354**, 136-139 (1991).
23. Macko, S. A. thesis, Univ. of Texas at Austin (1981).
24. Libes, S. M. thesis, Woods Hole/Massachusetts Inst. Technol. Joint Program in Oceanography (1983).
25. Altabet, M. A. & Francois, R. *Geochim. cosmochim. Acta* (submitted).
26. Wada, E. & Hattori, A. *Geomicrobiol. J.* **1**, 85-101 (1978).
27. Montoya, J. P. thesis, Harvard Univ. (1990).

28. Biggs, D. C. *et al. in Init. Rep. Ocean Drilling Program (Part A)* (eds Barker, P. F. & Kennett, J. P.) 77-86 (Ocean Drilling Program, College Station, Texas, 1988).
29. Cifuentes, L. A., Fogel, M. L., Pennock, J. R. & Sharp, J. H. *Geochim. cosmochim. Acta* **53**, 2713-2721 (1989).
30. Montoya, J. P., Horrigan, S. G. & McCarthy, J. J. *Geochim. cosmochim. Acta* **55**, 3627-3638 (1991).
31. Wada, E. & Hattori, A. *Geochim. cosmochim. Acta* **40**, 249-251 (1976).
32. Francois, R. & Altabet, M. A. *Paleoceanography* (in the press).
33. Boyle, E. A. & Lea, D. W. *Trans. Am. geophys. Un.* **70**, 1134 (1989).
34. Lohman, G. P. & Pride, C. J. *Trans. Am. geophys. Un.* **70**, 1134 (1989).
35. Stanley, D. J. *Nature* **274**, 149-152 (1978).
36. Rossignol-Strick, M. *Paleogeogr. Palaeoclim. Palaeoecol.* **49**, 237-263 (1985).
37. Fontugne, M. R. *et al. in Late Quaternary Chronology and Paleoclimates of the Eastern Mediterranean* (eds Bar-Yosef, O. & Kra, R. S.) (in the press).
38. Hoering, T. C. & Ford, H. T. *J. Am. chem. Soc.* **82**, 225-232 (1960).
39. Delwiche, C. C. & Steyn, P. L. *Envir. Sci. Technol.* **4**, 929-935 (1970).
40. Minagawa, M. & Wada, E. *Mar. Chem.* **19**, 245-259 (1986).
41. Checkley, D. M. Jr & Miller, C. A. *Deep-Sea Res.* **36**, 1449-1456 (1989).
42. Sournia, A. *Ann. Biol.* **9**, 63-76 (1970).
43. Zernova, V. V. *Tr. Inst. Okeanol. Akad. Nauk. SSSR* **58**, 45-53 (1962).
44. Carpenter, E. J. & Romans, K. *Science* **254**, 1356-1358 (1991).
45. Sarmiento, J. L., Herbert, T. & Toggweiler, J. R. *Global biogeochem. Cycles* **2**, 427-444 (1988).
46. Calvert, S. E. in *Marine Petroleum Source Rocks* (eds Brooks, J. & Fleet, A. J.) 137-151 (Geological Society of London, London, 1987).
47. Cita, M. B. *et al. Quat. Res.* **8**, 205-235 (1977).
48. Sweeney, R. E. & Kaplan, I. R. *Mar. Chem.* **9**, 81-94 (1980).
49. Rau, G. H., Arthur, M. A. & Dean, W. E. *Earth planet. Sci. Lett.* **82**, 269-279 (1987).

ACKNOWLEDGEMENTS. The collection of core MD 84641 was made possible by Territoire des Terres Australes et Antarctiques Françaises (NOE-Cruise of the R/V *Marion-Dufresne*). We thank A. Mariotti for preliminary $\delta^{15}\text{N}$ determinations, and M. Noyon and M. Soon for laboratory assistance. S.E.C. acknowledges support from the Natural Sciences and Engineering Research Council of Canada, the Canada Council, CNRS-CEA (France) and the Ministère Français des Affaires Étrangères.

Susceptibility of the early Earth to irreversible glaciation caused by carbon dioxide clouds

Ken Caldeira & James F. Kasting

Earth System Science Center & Department of Geosciences,
The Pennsylvania State University, University Park,
Pennsylvania 16802, USA

SIMPLE energy-balance climate models of the Budyko/Sellers type^{1,2} predict that a small (2–5%) decrease in solar output could result in runaway glaciation on the Earth. But solar fluxes 25–30% lower early in the Earth's history^{3,4} apparently did not lead to this result. One currently favoured explanation is that high partial pressures of carbon dioxide, caused by higher volcanic outgassing rates and/or slower rates of silicate weathering, created a large enough greenhouse effect to keep the planet warm^{5–7}. This does not resolve the problem of climate stability, however, because as we argue here, the oceans can freeze much more quickly than CO₂ can accumulate in the atmosphere. Had such a transient global glaciation occurred in the distant past when solar luminosity was low, it might have been irreversible because of the formation of highly reflective CO₂ clouds, similar to those encountered in climate simulations of early Mars⁸. Our simulations of the early Earth, incorporating the possible formation of such clouds, suggest that the Earth might not be habitable today had it not been warm during the first part of its history.

In the models of Budyko¹ and Sellers², a small reduction in solar luminosity results in the advance of ice and snow cover towards the Equator. This additional ice and snow reflects more solar radiation back to space and further cools the planet. If solar luminosity is reduced beyond some critical value in these models, this ice-albedo feedback leads to catastrophic freezing of Earth's entire surface.

Simple models like these cannot explain why the Earth was warm early in its history. The presence of sedimentary deposits in early Archaean rock sequences from Isua, West Greenland, shows that liquid water was present as early as 3.8 Gyr ago⁹, when solar luminosity was as much as 25% less, according to standard solar evolution models³. High concentrations of greenhouse gases, such as ammonia⁴ or carbon dioxide⁵ can explain the warm climate; the problem then is to explain why the greenhouse gases were abundant. Walker *et al.*⁶ proposed that

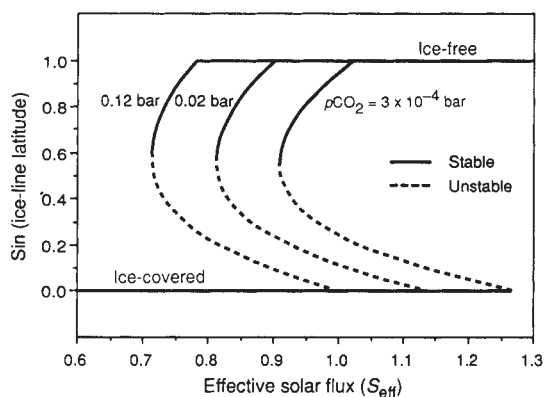


FIG. 1 Steady-state ice lines (x_s) as a function of effective solar luminosity, S_{eff} , for three values of atmospheric $p\text{CO}_2$. If a perturbation were to shift the Earth from its present state ($S_{\text{eff}}=1$, $p\text{CO}_2=3 \times 10^{-4}$ bar, $x_s=0.95$) to an ice-covered state today ($S_{\text{eff}}=1$, $p\text{CO}_2=3 \times 10^{-4}$ bar, $x_s=0$), then sufficient CO₂ (~0.12 bar) would accumulate in the atmosphere within 30 Myr to make the ice-covered state unstable. The model would then shift to the ice-free state ($S_{\text{eff}}=1$, $p\text{CO}_2=0.12$ bar, $x_s=1$) and silicate-rock weathering would begin to remove the excess CO₂ from the atmosphere.

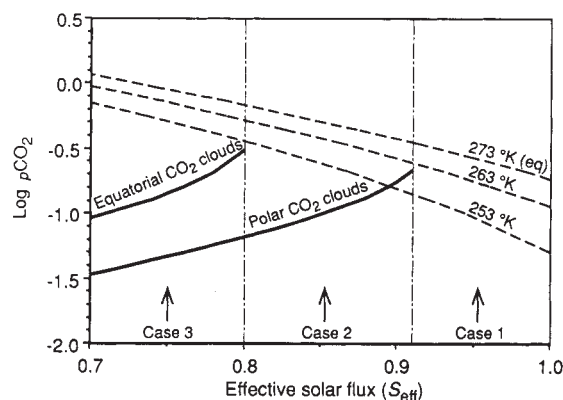


FIG. 2 Solid curves indicate the onset of polar and global CO₂-cloud cover, dashed curves show equatorial surface temperature as a function of effective solar luminosity and atmospheric $p\text{CO}_2$. Case 1: starting from an ice-covered state with low atmospheric $p\text{CO}_2$, volcanic CO₂ would accumulate in the atmosphere and initiate ice-melting before the formation of CO₂ clouds. Case 2: polar CO₂ clouds would form before the onset of equatorial ice-melting. Case 3: CO₂ clouds would form globally before the onset of ice-melting, in which case CO₂-global warming would probably be incapable of melting off the ice.

the temperature dependence of the silicate weathering rate creates a negative feedback that maintains high enough CO₂ concentrations to stabilize the Earth against the ice-albedo catastrophe. This negative feedback was incorporated into an energy-balance climate model by Marshall *et al.*⁷, who showed that it could produce cold global climates if the continents were clustered near the Equator. Exposed silicate rock would then be bathed in a relatively warm and wet environment, conducive to CO₂ consumption by weathering. Both lower solar luminosity and the presence of an equatorial, Late Proterozoic supercontinent may therefore help to explain the widespread low-latitude Late Proterozoic glaciation^{7,10–13}.

Marshall *et al.*⁷ further suggested that the silicate-weathering/CO₂ feedback can stabilize the ice line at low latitudes, but reanalysis of their model indicates that this is incorrect. Because atmospheric CO₂ responds much more slowly ($>10^5$ yr; ref. 14) than does sea ice and snow cover (<1 yr), the silicate-weathering feedback cannot buffer the climate system against rapid fluctuations in the ice line. A perturbation analysis¹⁵ of their model shows that there is no stable ice line nearer to the Equator than $\sim 30^\circ$ latitude, regardless of the partial pressure of atmospheric CO₂. (The convergence problems mentioned by Marshall *et al.*⁷ may have been a symptom of this instability.) In their model, a low-latitude glaciation should therefore run away to the globally ice-covered state. Could this have actually happened during the Late Proterozoic era?

To study this question, and to address the general problem of climate stability throughout geological time, we developed a new zonally averaged energy-balance climate model (see Box 1 for description). A key difference between our model and previous energy-balance models is that we have taken into account the possible formation of CO₂-ice clouds. Such clouds may have formed in the atmosphere of early Mars⁸ and could have formed on an ice-covered early Earth as well.

At today's solar flux and CO₂ level, our model shows four possible steady states, where x_s is the sine of the ice-line latitude (Fig. 1): ice-free ($x_s=1$), stable partial ice cover ($x_s=0.95$), unstable partial ice cover ($x_s=0.28$) and ice-covered ($x_s=0$). If the Earth was initially in the stable, partially ice-covered state, and was then subjected to a rapid perturbation (relative to the response time of the silicate weathering feedback) that would either temporarily lower the effective solar flux to about 0.9, or produce transient glaciation equatorward of $x_s=0.28$, the Earth would fall into the ice-covered state ($x_s=0$). Without any change

BOX 1 Energy-balance climate model

Our zonally averaged energy-balance climate model uses albedo and outgoing infrared parameterizations derived from a one-dimensional radiative-convective climate model^{8,17,20}. In the energy-balance model, the rate of solar energy input to a latitude belt is locally balanced by the sum of the energy leaving the latitude belt as infrared radiation to space and the net heat transport to other latitude belts. This may be represented mathematically by the relation²¹

$$-\frac{d}{dx} D(1-x^2) \frac{dT(x)}{dx} + I(x) = S(x)(1-a_{\text{toa}}(x)) \quad (1)$$

where D is a meridional heat diffusion coefficient, x is the sine of latitude, $T(x)$ is the zonally averaged temperature in a given latitude band, $I(x)$ is the outgoing infrared radiation to space, $S(x)$ is the annual radiation reaching the top of the atmosphere and $a_{\text{toa}}(x)$ is the zonally averaged top-of-atmosphere albedo. Boundary conditions on equation (1) are that meridional heat transport vanishes at both the Equator and the poles. Average mean solar energy incident at a given latitude is represented as $S(x) = Q(1 - 0.477P_2(x))$ (ref. 22), where Q is the solar luminosity at the Earth's orbital distance ($=S_{\text{eff}} \times 1,360 \text{ W m}^{-2}$), and $P_2(x) = (3x^2 - 1)/2$ is the second Legendre polynomial.

Outgoing infrared radiation, $I(x)$, may be parameterized as a function of surface temperature by an equation of the form^{1,2}

$$I(x) = A + BT(x) \quad (2)$$

The coefficients A and B depend on the atmospheric CO_2 partial pressure⁷. We determined these coefficients by performing least-squares fits, with an r.m.s. error under 5 W m^{-2} , to more than 2,000 runs of the radiative-convective climate model^{8,17,20}, for $10^{-4} \text{ bar} < P < 2 \text{ bar}$, where P is CO_2 partial pressure, and $194 \text{ K} < T < 303 \text{ K}$. The fitting process yielded

$$A(\varphi) = -326.4 + 9.161\varphi - 3.164\varphi^2 + 0.5468\varphi^3 \quad (3)$$

and

$$B(\varphi) = 1.953 - 0.04866\varphi + 0.01309\varphi^2 - 0.002577\varphi^3 \quad (4)$$

where φ is the natural log of atmospheric $p\text{CO}_2$ divided by a reference level of 300 p.p.m. In the radiative-convective model^{8,17,20}, the troposphere is considered to be fully saturated with H_2O . Assuming a lower relative humidity would make the model even more susceptible to irrevers-

ible glaciation. The effect of clouds on the outgoing infrared flux was computed by subtracting a constant amount (15.56 W m^{-2}) from the cloud-free value. This allows the model to reproduce the present mean surface temperature, given the observed solar insolation and planetary albedo. Holding the cloud infrared constant at lower surface temperatures should again produce an upper limit on greenhouse warming, as cloud cover could have been reduced under such circumstances²³.

The top-of-atmosphere albedo (a_{toa}) at a given latitude is a function of surface albedo (a_s), solar zenith angle (θ) and the abundances of CO_2 , N_2 and H_2O . We used the radiative-convective model to develop a polynomial representation of a_{toa} as a function of these quantities

$$\begin{aligned} a_{\text{toa}} = & -0.6711 + 1.573a_s - 0.01988P + 0.005749T + 0.03815\mu \\ & - 0.1003a_sP - 0.002478a_sT - 0.0006519a_s\mu \\ & + 0.0001603PT - 0.0001957P\mu - 0.0001123T\mu \\ & - 0.1846a_s^2 + 0.01747P^2 - 0.00001143T^2 + 0.009790\mu^2 \end{aligned} \quad (5)$$

Here, $P = p\text{CO}_2$ in bars and $\mu = \cos \theta$. The value of μ in a given latitude band is $S(x)/(2Q)$. We calculated the effective surface albedo by finding the value needed to produce the latitudinal distribution of planetary albedo²², obtaining $a_s(x) = 0.2595 + 0.210P_2(x)$, for ice-free surfaces. This expression includes the effect of (present-day) clouds. For areas with mean annual temperature below -10°C , the surface was assumed to be ice-covered and to have a surface albedo of 0.663; this gives a planetary albedo of 0.62 under modern polar conditions. Following others^{7,21}, we set the diffusion coefficient ($D = 0.6334 \text{ W m}^{-2} \text{ K}^{-1}$) to produce a present-day ice line (x_0) equal to 0.95.

The model forms CO_2 clouds at sine latitude x when

$$T(x) < 281 + 68.2\psi + 15.4\psi^2 + 1.34\psi^3 \quad (6)$$

where ψ is the common log of surface $p\text{CO}_2$ in bars (Fig. 3). This inequality represents the radiative convective model results to within 2 K when $-5 < \psi < 1$.

$T(x)$ itself was approximated as $T(x) = T_0 + T_2P_2(x)$ (ref. 21). Numerical experiments showed that retaining an additional T_4 term did not significantly alter our results.

in atmospheric $p\text{CO}_2$, an increase in solar flux by $\sim 27\%$ above the present value would be needed to melt the equatorial ice (Fig. 1). At solar fluxes higher than this, a reverse ice-albedo feedback would completely deglaciate the Earth. But this is not what would actually happen were the present Earth to freeze. In the ice-covered state little or no silicate rock would be exposed to weathering, so CO_2 from metamorphic and mantle sources could accumulate in the atmosphere at a rate of $\sim 8 \times 10^{12} \text{ mol yr}^{-1}$ (ref. 16). In less than 30 Myr, atmospheric $p\text{CO}_2$ would build up to nearly 0.12 bar, and equatorial ice would become unstable (Fig. 2). This evolution corresponds to case 1 in Fig. 3.

Suppose now that this same perturbation, from stable partial

ice cover to global glaciation, occurred earlier in Earth's history when solar luminosity was lower. Our climate model indicates that the results could then be very different. Colder temperatures would permit the formation of CO_2 ice clouds in the upper troposphere. These clouds would further cool the surface by reducing the tropospheric lapse rate (because of the release of latent heat) and by reflecting the additional sunlight back to space. The first of these processes is simulated in our model by assuming that the tropospheric lapse rate follows a moist CO_2 adiabat in the region where CO_2 reaches saturation^{8,17}. The second process is not included in our numerical calculations because of uncertainties regarding the horizontal extent and optical depth of these clouds and because doing so would require

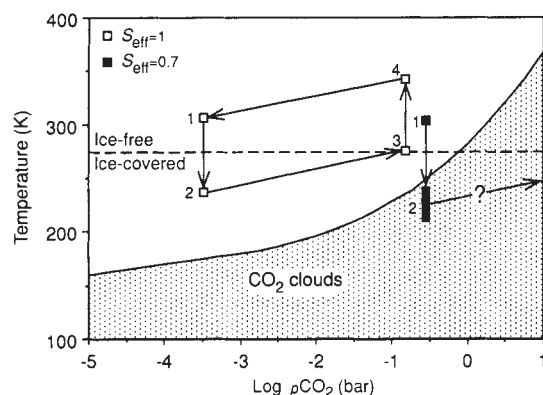


FIG. 3 Evolution of equatorial surface temperature in the event of catastrophic glaciation for two values of effective solar luminosity, S_{eff} . $S_{\text{eff}} = 1$: if a perturbation were to shift the Earth from its present state (open square 1) to an ice-covered state (open square 2), the Earth's surface would be too warm for CO_2 clouds to form. Carbon dioxide could therefore accumulate in the atmosphere, warming the Earth. The ice-covered state would become unstable (open square 3), and the Earth would shift to the ice-free state (open square 4). Silicate-rock weathering could then remove the excess CO_2 from the atmosphere, returning the system to its present state (open square 1). $S_{\text{eff}} = 0.7$: if a perturbation were to shift the Earth from an ice-free state at $p\text{CO}_2 = 0.25 \text{ bar}$ (filled rectangle 1) to an ice-covered state (filled rectangle 2), the Earth's surface would be cold enough for CO_2 clouds to form. These clouds would have a cooling effect, and increasing atmospheric $p\text{CO}_2$ may not warm the Earth enough to escape this CO_2 -cloud covered state. The upper bound of rectangle 2 represents the calculated equatorial temperature neglecting the cooling effect of the CO_2 clouds. The lower bound represents the equatorial temperature if the CO_2 clouds produced a planetary albedo of 0.8. In this case, the Earth would remain covered with CO_2 clouds and surficial water ice even if solar luminosity were increased to today's value ($S_{\text{eff}} = 1$).

a radiative model that incorporates scattering in the infrared. (CO₂ ice, unlike water or water ice, is a nearly pure scatterer at most infrared wavelengths¹⁸.) The radiative effect of CO₂ clouds can, however, be estimated qualitatively: because CO₂ clouds are poor absorbers, their contribution to the greenhouse effect should be relatively small, so their primary influence should be to cool the Earth by increasing its albedo.

The points where CO₂ clouds begin to form in our model are indicated by the solid curves in Fig. 2. We have divided the earlier phase of the Earth's history into two parts, labelled 'case 2' and 'case 3'. When S_{eff} is less than ~ 0.92 (earlier than ~ 1 Gyr ago³) our model predicts that CO₂ clouds would start to form at the poles (case 2). When S_{eff} is less than ~ 0.8 (earlier than ~ 3 Gyr ago³) CO₂ cloud cover would extend all the way down to the Equator. It is uncertain what the albedo of such a CO₂ cloud-covered planet would be. If it were as high as 0.8, our model predicts that the Earth would not be able to emerge from that state, even at the present solar luminosity. Thus, if the Earth had experienced runaway glaciation before ~ 3 Gyr ago, the situation might have been effectively irreversible.

How the early Earth managed to avoid becoming globally glaciated is a question of considerable importance to those interested in the general problem of planetary habitability. One possibility is that the Earth remained in a low-ice state simply because it started out hot following accretion: a dense, CO₂-N₂-H₂O atmosphere that was initially warm would be free of CO₂ clouds even at 70% of present solar luminosity (Fig. 3). Alterna-

tively, the presence of additional greenhouse gases, such as NH₃ (ref. 4) or CH₄ (ref. 19), might have kept the atmosphere warm enough to prevent CO₂ from condensing. In either case, the explanation for why Earth's climate has remained stable is more complex than has been previously assumed. □

Received 24 April; accepted 28 July 1992.

1. Budyko, M. I. *Tellus* **21**, 611-619 (1969).
2. Sellers, W. D. *J. appl. Met.* **8**, 392-400 (1969).
3. Gough, D. O. *Solar Phys.* **74**, 21-34 (1981).
4. Sagan, C. & Mullen, G. *Science* **177**, 52-56 (1972).
5. Owen, T., Cess, R. D. & Ramanathan, V. *Nature* **277**, 640-642 (1979).
6. Walker, J. C. G., Hays, P. B. & Kasting, J. F. *J. geophys. Res.* **86**, 9776-9782 (1981).
7. Marshall, H. G., Walker, J. C. G. & Kuhn, W. R. *J. geophys. Res.* **93**, 791-801 (1988).
8. Kasting, J. F. *Icarus* **94**, 1-13 (1991).
9. Allart, J. H., in *The Early History of the Earth* (ed. Windley, B. F.) 177-189 (Wiley, New York, 1976).
10. Worsley, T. R. & Kidder, D. L. *Geology* **19**, 1161-1164 (1992).
11. Walter, M. *Am. Scientist* **67**, 142 (1979).
12. McWilliams, M. O. & McElhinney, M. W. *J. Geol.* **88**, 1-26 (1980).
13. Hambrey, M. J. & Harland, W. B. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **51**, 255-272 (1985).
14. Sundquist, E. T. *Quat. Sci. Rev.* **10**, 283-296 (1991).
15. Cahalan, R. F. & North, G. R. *J. Atmos. Sci.* **36**, 1178-1188 (1979).
16. Holland, H. D. *The Chemistry of the Atmosphere and Oceans* (Wiley, New York, 1978).
17. Kasting, J. F., Whitmire, D. P. & Reynolds, R. T. *Icarus* (in the press).
18. Warren, S. G. *Appl. Opt.* **25**, 2650-2674 (1986).
19. Kiehl, J. T. & Dickinson, R. E. *J. geophys. Res.* **92**, 2991-2998 (1986).
20. Kasting, J. F. & Ackerman, T. P. *Science* **234**, 1383-1385 (1986).
21. North, G. R., Cahalan, R. F. & Coakley, J. A. *Rev. Geophys.* **19**, 91-121 (1981).
22. North, G. R. & Coakley, J. A. *J. Atmos. Sci.* **36**, 1189-1204 (1979).
23. Rossow, W. B., Henderson-Sellers, A. & Weinrich, S. K. *Science* **217**, 1245-1247 (1982).

ACKNOWLEDGEMENTS. We thank A. Lapienis for discussions and C. Sagan for a review. K.C. was supported by the NSF under a grant awarded in 1991.

Role of pore fluids in the generation of seismic precursors to shear fracture

P. R. Sammonds*, P. G. Meredith* & I. G. Main†

* Department of Geological Sciences, University College London, Gower Street, London WC1E 6BT, UK

† Department of Geology & Geophysics, University of Edinburgh, Grant Institute, West Mains Road, Edinburgh EH9 3JW, UK

A SYSTEMATIC study of temporal changes in seismic b -values (defined as the log-linear slope of the earthquake frequency-magnitude distribution) has shown that large earthquakes are often preceded by an intermediate-term increase in b , followed by a decrease in the months to weeks before the earthquake¹. The onset of the b -value increase can precede earthquake occurrence by as much as 7 years. A recently proposed fracture mechanics model of the earthquake source² explains these temporal fluctuations in b in terms of the underlying physical processes of time-varying applied stress and crack growth. The model predicts two minima in b , separated by a short-lived maximum. Here we report the results of controlled laboratory deformation experiments, done in simulated upper-crustal conditions on both air-dried and water-saturated rock specimens. As found in previous experiments³⁻⁵, shear fracture in dry specimens is characterized by a decline in b during anelastic deformation to a single minimum reached just before failure. But in water-saturated specimens, when pore-fluid volume is kept constant by servo-control we also observe a second, intermediate-term b -value minimum, so reproducing the double b -value anomaly predicted by the model².

The fracture mechanics model of Main *et al.*² is summarized in Fig. 1. In nature, because crustal deformation is accompanied and accomplished by microcracking and small earthquakes⁶, sometimes coupled with aseismic deformation, the crust will respond to remotely applied tectonic loading by deforming anelastically; it shows both strain-hardening and strain-

softening phases, and a dynamic failure stress lower than peak stress (Fig. 1a). Laboratory experiments have shown that a principal flaw under the action of remotely applied stress and stress corrosion processes will extend in a strongly nonlinear fashion⁷ (Fig. 1b) from an initial length x_0 at time t_0 to effectively infinite length at time t_f . Combining the effects of stress and

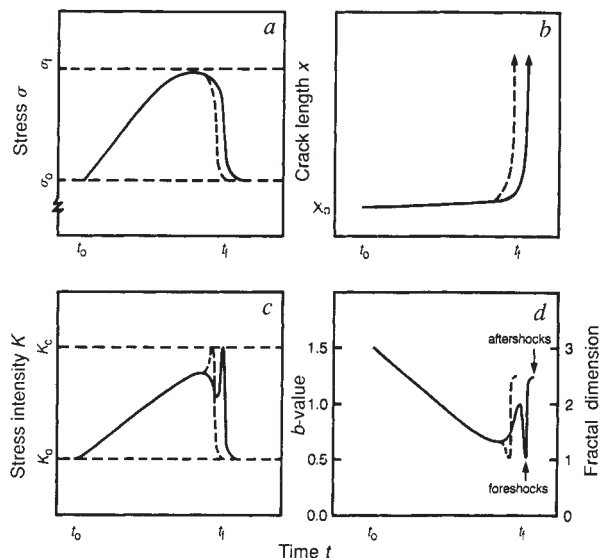


FIG. 1 Fracture mechanics model of the b -value anomaly for anelastic failure (precursory strain energy release model)². a, Stress-time behaviour for rock deformation; b, acceleration of crack tip to critical failure; c, combined effects of the stress and crack length of the stress intensity factor, K . d, Because b is negatively correlated and linearly related to K , the model predicts both an intermediate-term and a short-term b -value minimum. A single minimum will be produced where failure occurs during an earlier part of the cycle (dashed line). σ_0 , initial stress at time t_0 ; σ_f , nominal failure stress (peak stress); t_0 , onset time of loading; t_f , dynamic failure time; x_0 , crack length at time t_0 ; K_0 , critical stress intensity (fracture toughness); K_0 , stress intensity at time t_0 .

crack growth behaviour, the stress intensity factor K (defined by $K = Y\sigma x^{1/2}$, where σ is the remotely applied stress and Y is a geometric factor) is modelled as increasing gradually from a value K_0 at time t_0 to a level approaching a critical value for macroscopic failure K_c (Fig. 1c). But decreasing stress in the strain-softening phase reduces K temporarily before the effect of the accelerating crack length begins to dominate, resulting in a rapid increase in K to critical failure. As it has been found

experimentally that the stress intensity factor of the principal flaw is negatively correlated to the b -value derived from monitoring acoustic emissions produced by microcracking as the flaw advances⁸, the model predicts both an intermediate-term and short-term b -value anomaly in the form of a double minimum, associated with K reaching a maximum first at peak stress and then at the onset of dynamic failure (Fig. 1d). A single minimum in b will, however, be produced where failure occurs during an

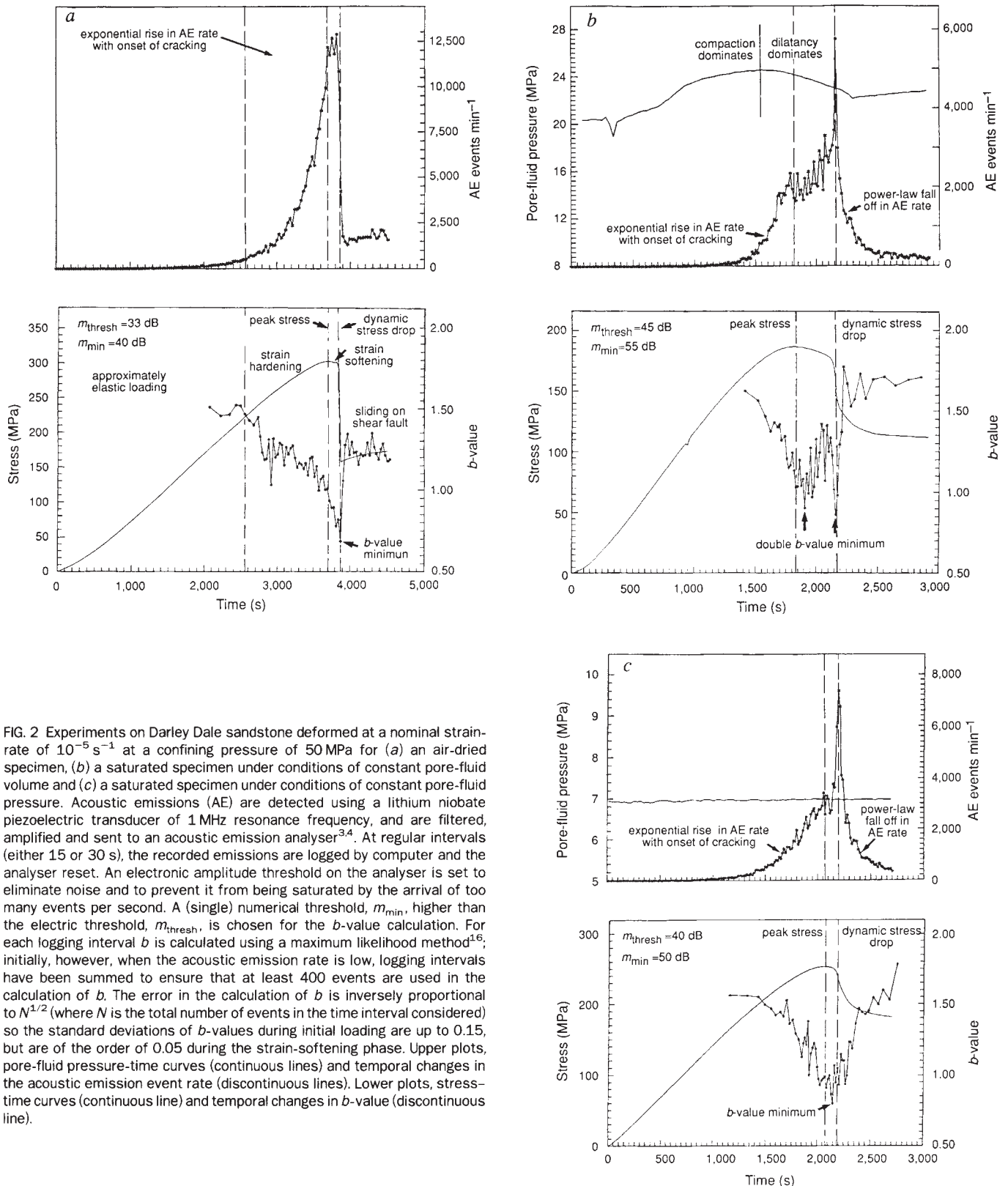


FIG. 2 Experiments on Darley Dale sandstone deformed at a nominal strain-rate of 10^{-5} s^{-1} at a confining pressure of 50 MPa for (a) an air-dried specimen, (b) a saturated specimen under conditions of constant pore-fluid volume and (c) a saturated specimen under conditions of constant pore-fluid pressure. Acoustic emissions (AE) are detected using a lithium niobate piezoelectric transducer of 1 MHz resonance frequency, and are filtered, amplified and sent to an acoustic emission analyser^{3,4}. At regular intervals (either 15 or 30 s), the recorded emissions are logged by computer and the analyser reset. An electronic amplitude threshold on the analyser is set to eliminate noise and to prevent it from being saturated by the arrival of too many events per second. A (single) numerical threshold, $m_{\min}$, higher than the electric threshold, m_{thresh} , is chosen for the b -value calculation. For each logging interval b is calculated using a maximum likelihood method¹⁶; initially, however, when the acoustic emission rate is low, logging intervals have been summed to ensure that at least 400 events are used in the calculation of b . The error in the calculation of b is inversely proportional to $N^{1/2}$ (where N is the total number of events in the time interval considered) so the standard deviations of b -values during initial loading are up to 0.15, but are of the order of 0.05 during the strain-softening phase. Upper plots, pore-fluid pressure-time curves (continuous lines) and temporal changes in the acoustic emission event rate (discontinuous lines). Lower plots, stress-time curves (continuous line) and temporal changes in b -value (discontinuous line).

earlier part of the cycle (dashed lines in Fig. 1). Calculations of t_f for conditions found on the San Andreas fault suggest rupture times of 150 to 220 years, but with sudden crack length changes which give rise to seismic precursors occurring only in the final 10 years⁹.

To test this model we have done triaxial deformation experiments on initially intact specimens of Darley Dale sandstone, a well-indurated felspathic sandstone with a siliceous cementing material, selected because it shows distinct phases of both strain hardening and strain softening during deformation. Right cylindrical specimens 15 mm in diameter by 45 mm in length, jacketed in a ductile copper sleeve, were deformed in compression in a high-pressure triaxial cell, incorporating a servo-controlled pore-fluid pressure intensifier and volume monitor. An all-round hydrostatic pressure is first applied to the specimen and maintained at a set value (the 'confining pressure'). The pore-fluid intensifier piston is advanced to force high-pressure water into the specimen and left to equilibrate. Then an axial load is applied to the rock specimen by a 200-kN servo-controlled actuator at constant displacement rate. During deformation of the specimen the pore-fluid intensifier can be set to maintain either a constant pore-fluid pressure or constant pore-fluid volume. Acoustic emission is detected continuously using a transducer located close to the specimen in the hollow lower loading piston inside the pressure vessel.

Figure 2a shows acoustic emission results from an experiment on an air-dried sandstone specimen. Initial compaction of the specimen is associated with low acoustic emission. At ~50% of peak stress, microcrack initiation and extension commences and the rate of acoustic emission activity increases exponentially. Peak stress is followed by a small but significant period of strain softening leading to dynamic failure. Failure occurs on a well-defined fault plane oriented at between 30° and 35° to the maximum principal compression axis. During the quasi-linear elastic phase of loading, the b -value is high, fluctuating around an average of ~1.5. As deformation proceeds, b decreases steadily; then just before dynamic failure, it dips sharply towards ~0.5. It then recovers to ~1.2 during the period of frictional sliding that follows failure. These experimentally measured b -values roughly agree with those measured in nature, with background seismicity often associated with $b \geq 1$ and earthquake foreshocks with $b = 0.5$ (ref. 10), and with experimental results presented previously³⁻⁵. Observations of the intermediate-term increase in b followed by a decrease were not, however, reproduced.

By contrast, Fig. 2b shows results from an experiment where the specimen was deformed in the presence of water as a pore fluid held at constant pore-fluid volume. (These conditions roughly correspond to a depth of 2-3 km with a pore-fluid pressure that is initially hydrostatic.) On comparison with Fig. 2a, the stress-time curve shows that the specimen is considerably weaker because of the combined chemical and mechanical influences of the pressurized pore fluid, and that dynamic failure is now preceded by a prolonged period of strain softening due to dilatancy hardening caused by a decreasing pore-fluid pressure. After a period of stability during initial loading, the pore-fluid pressure starts to increase as the specimen is compacted, peaking as the opening of new dilatational microcracks begins to dominate over the closure of pre-existing cracks and flaws, and then falls, finally reaching an essentially constant level during the frictional sliding phase. The acoustic emission rate increases roughly exponentially with time, reaching an inflection point around peak stress. After failure the rate decreases roughly according to a power law, consistent with observations of after-shock sequences¹¹. But it is the b -value that show the most interesting behaviour. The experiment again reproduces the now-familiar short-term b -value minimum, but we now observe an intermediate-term b -value minimum, followed by a recovery before dipping to the second minimum just before the dynamic stress drop. Of all the conditions under which we have deformed

rocks, this test condition of shear fracture at controlled constant pore-fluid volume is the only one to show the distinctive 'double minimum' in b . The recovery in b and short-term drop correspond to the type of anomaly reported by Smith¹.

Finally, in Fig. 2c we present results from an experiment done under conditions of constant pore-fluid pressure. The application of even a small constant pore-fluid pressure to the specimen results in a stress-time curve that shows only a short strain-softening phase and small dynamic stress drop. A single b -value minimum is observed just before the dynamic failure under these conditions.

Our experimental results reveal an exponentially increasing rate of acoustic emission during loading under all three conditions up to peak stress, but for the experiment at constant pore-fluid volume there is abrupt flattening of the rate. This last case we interpret as being caused by local dilatancy hardening resulting from falling pore-water pressure, although the specimen is softening overall. This has the effect of extending the post-peak stress deformation before dynamic failure occurs. A declining b -value is observed initially for anelastic deformation for all three test conditions. For the experiments with air-dry specimens or constant pore-fluid pressure, the decrease in b leads directly to dynamic failure immediately following minimum b . Even though overall stress was falling as the specimen strain-softened, this was not sufficient to relieve the stress intensity driving these systems to failure. For the experiment at constant pore-fluid volume, the b -value first recovered before falling to a second minimum at failure. In this last case, the combined effects of falling applied stress and falling pore-fluid pressure allowed a relaxation in stress intensity, with damage accumulation still progressing but with the formation of the shear fault delayed.

The earthquake source model of Main *et al.*² requires a prolonged strain-softening phase before dynamic failure if a double b -value minimum is to be observed in nature. This could be promoted by a falling pore-fluid pressure at constant pore-fluid volume. (The model also accounts for b -value behaviour for a short strain-softening phase). Nonhydrostatic pore fluid pressures in the crust require that rock permeability is low or that low-permeability rock encloses rock of higher permeability. There is considerable geological and geophysical evidence for this^{12,13}, and time-dependent changes in permeability in the crust have been roughly quantified¹². The departure of pore-fluid pressure from hydrostatic has also been cited as an important mechanism in the 'dilatancy-diffusion' model of the earthquake source¹⁴ and to account for the 'weakness' of the San Andreas fault¹⁵. Extrapolation of recent calculations¹² suggest that small changes in pore-fluid pressure needed to prolong the strain softening phase could occur within 10 years, but a full quantitative description of the model of Main *et al.*² awaits an understanding of crack development, porosity and permeability changes for the complete deformation cycle. □

Received 4 March; accepted 3 August 1992.

- Smith, W. D. *Nature* **289**, 136-139 (1981).
- Main, I. G., Meredith, P. G. & Jones, C. *Geophys. J.* **96**, 131-138 (1989).
- Meredith, P. G., Main, I. G. & Jones, C. *Tectonophysics* **175**, 249-268 (1990).
- Sammonds, P. R., Ayling, M. R., Meredith, P. G., Murrell, S. A. F. & Jones, C. in *Rock at Great Depth* (eds Maury, V. & Fourmairiaux, D.) (Balkema, Rotterdam, 1989).
- Lockner, D. A., Byerlee, J. D., Kiksenko, V., Ponomarev, A. & Sidorin, A. *Nature* **350**, 39-42 (1991).
- Mogi, K. *Earthquake Prediction* (Academic, London, 1985).
- Atkinson, B. K. & Meredith, P. G. in *Fracture Mechanics of Rock* (ed. Atkinson, B. K.) (Academic, London, 1987).
- Meredith, P. G. & Atkinson, B. K. *Geophys. J. R. astr. Soc.* **75**, 1-21 (1983).
- Main, I. *Geophys. J.* **92**, 455-464 (1988).
- von Seggern, D. *Geophys. J. R. astr. Soc.* **86**, 815-838 (1980).
- Utsu, T. *Geophys. Mag.* **30**, 521-605 (1961).
- Nur, A. M. & Walder, J. in *The Role of Fluids in Crustal Processes* (panel co-chair Bredehoeft, J. D. & Norton, D. L.) (National Academy Press, Washington DC, 1990).
- Marquis, G. & Hyndman, R. D. *Geophys. J. int.* **110**, 91-105 (1992).
- Scholz, C. H., Sykes, L. R. & Aggarwal, Y. P. *Science* **181**, 803-810 (1973).
- Rice, J. R. *Extended Abstr. int. Symp. Earthq. Source Phys. Earthq. Precursors 7-11* (Univ. of Tokyo, 1990).
- Aki, K. *Bull. Earthq. Res. Inst. Tokyo Univ.* **43**, 237-239 (1965).

ACKNOWLEDGEMENTS. Financial support for this work was provided by the U.K. NERC.

Surfaces versus features in visual search

Zijiang J. He & Ken Nakayama

Vision Sciences Laboratory, Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, Massachusetts 02138, USA

OFTEN implicit in the interpretation of visual search tasks is the assumption that the detection of targets is determined by the feature-coding properties of low-level visual processing^{1,2}. But higher level processes have also been implicated as visual search ability is enhanced in a depth plane³ or when two-dimensional shapes are interpreted as three-dimensional forms^{4,5}. Here we manipulate binocular disparity to degrade visual search, so that otherwise identical features become parts of surfaces through perceptual completion, rendering them less clearly distinguishable as targets and distractors. Our results indicate that visual search has little or no access to the processing level of feature extraction but must have as an input a higher level process of surface representation.

Stereoscopic depth can have a profound effect on the interpretation of surface continuation behind occluders⁶ and can control perceived transparency⁷. For example, consider what happens to an L-shaped target when binocular disparity is manipulated (see stereograms, Fig. 1). In Fig. 1a, the L shape is seen in front, standing alone and appearing to cover the adjacent square. In Fig. 1b, the L-shape is seen behind, but no longer appears as

an L-shape, instead taking on the appearance of a square continuing behind the other⁶. The strong perceptual continuation of a surface behind an occluder, referred to as 'amodal' completion⁸ has been shown to influence a wide variety of seemingly low-level visual functions, in particular motion perception^{9,10} and image segmentation⁶.

A similar question may be asked regarding the underlying mechanisms of visual search, where an observer must find an odd target among distracting elements. If visual search is dictated by the properties of features, as has been generally assumed, then subtle manipulations that leave these features intact should not influence performance. Alternatively, if visual search has little or no access to visual features, then such manipulations as seen in Fig. 1 should have a great effect. When targets and distractors are perceived as surfaces behind an occluder and thus no longer seen as Ls and reversed Ls, the search for the odd target should become much more difficult.

The various stimulus elements used in our first experiment are illustrated in Fig. 1a and b. The search stimulus consisted of one target (L-shape or mirror-reversed L-shape) and several distractors (L or reversed L) accompanied by squares. The target and distractor were always in a same depth plane, but appeared to be in a different depth plane relative to the squares. Seven observers (three naive) were used. Observers viewed the stereograms using a phase haploscope device, consisting of shuttered glasses to synchronize transmission in each eye with alternating TV monitor frames. The task was to search for a target among randomly distributed distractors and to release a button when the target was found. Responses in naive observers

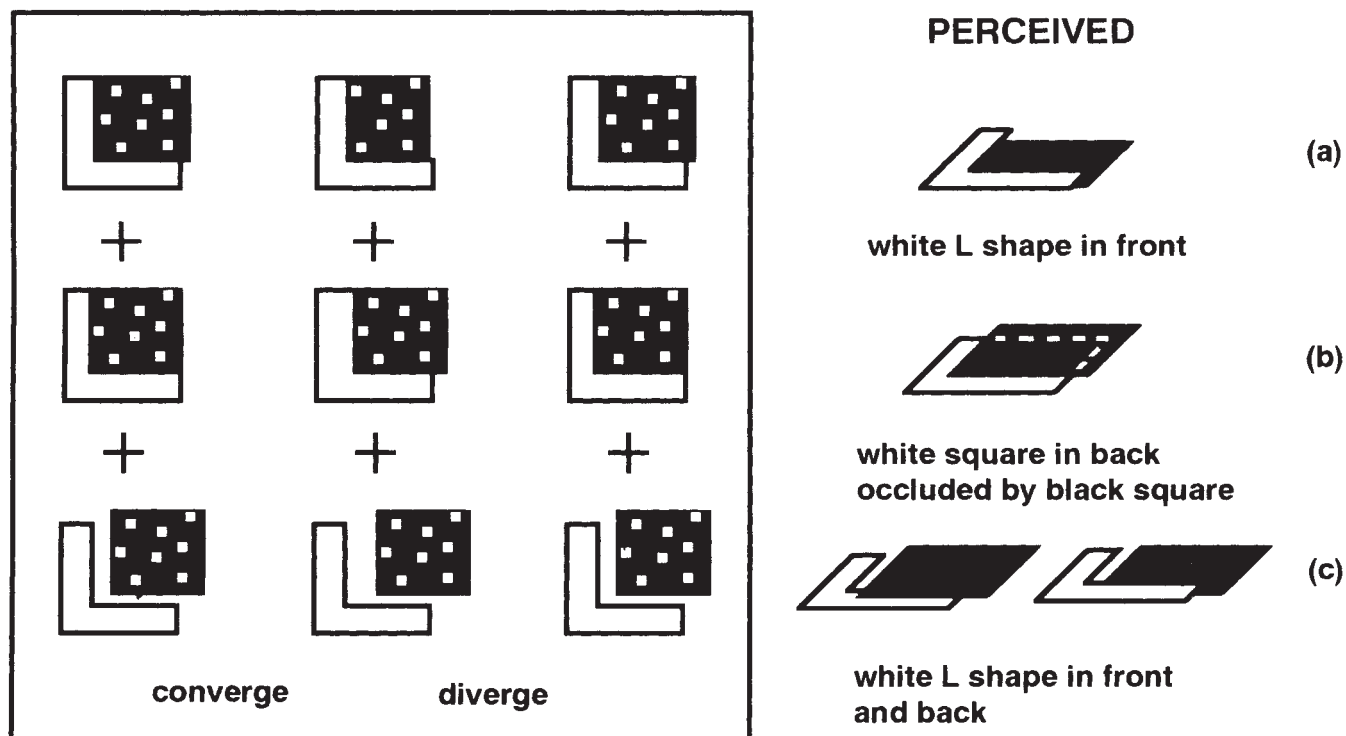


FIG. 1 Three image stereograms (in box) designed for convergent or divergent fusion. For convergent fusion, left and centre images should be used, ignoring the right image. For divergent fusion, use centre and right images. Perceived shapes are illustrated to the right. An L-shape is seen in front for row a, and behind for row b. In row b, the L becomes extended to form a larger surface, 'amodally completing behind the square occluder. Row c, stimuli used in the second experiment, which have the same disparity

differences but with a 9.2' separation between L and square. In both experiments, target and distractors were distributed randomly within a rectangular area (13.4° × 9.6°). The luminance of the grey background was 1.7 cd m⁻², the white L, 4.8 cd m⁻² and the black squares 0.6 cd m⁻². The size of the L was 1° × 1°, with the horizontal limb subtending 11.5'. Crosses are provided to aid in free fusion.

TABLE 1 Mean reaction time differences between presentation of L-shape behind and L-shape in front of square (ms)

Observer	Experiment 1*	Experiment 2
B.A.	1,154	-90
P.V.	1,083	63†
Q.W.‡	1,434	79
S.S.‡	752	-35
T.K.‡	1,253	48
T.W.	1,270	189†
Z.H.	2,279	107
Average	1,316 ± 474	52 ± 92

* The reaction time differences of each observer were significantly different (*t*-test; $P < 0.01$).

† Reaction time differences significantly different (*t*-test; $P < 0.05$).

‡ Naive observer.

were verified by requiring them to indicate whether the target appeared on the left or right half of the screen. In all experiments, location errors never exceeded 10% and had a mean of about 5%. Reaction times corresponding to incorrect position reports were not counted in final results.

Reaction time histograms for observer P.V. are shown in Fig. 2a. There is a significant shift in the reaction time distribution between the presentation of the target behind the square (upper histogram) and in front of the square (lower histogram). The other observers gave very similar results (Table 1, first column). For seven observers, the mean reaction time differences between target behind and target in front ranged from 752 ms to 2,279 ms, with an average of 1,316 ms.

Our results support the view that the increased reaction time when the L-shape is behind is caused by the perceptual or amodal⁸ completion of a putative surface behind the square, because perceptual completion is possible only in this case. But it is important to rule out the possibility that detection of targets might simply be more effective if they are in a front disparity plane¹¹.

To test this possibility we conducted a second experiment, comparing the detection of L-shapes in front of and behind the square, as in the previous experiments, but where there was no opportunity for perceptual completion behind the square. This was accomplished by creating a small gap between the square and L-shape, such that they did not share a common border (Fig. 1c). The manipulation of binocular disparity no longer provides the opportunity for amodal surface completion¹⁰. In both cases, the L-shapes appear as distinct entities. If our hypothesis regarding the role of surface completion is correct, then reversing binocular disparity should not lead to an appreciable difference in search reaction times for this experiment.

Indeed we found that there was very little difference between the presentation of L-shape in front and behind, when the Ls did not share a common border with the squares. Observer P.V.'s

reaction time histograms for this experiment are shown in Fig. 2b. Table 1 (second column) shows the difference between mean reaction times (behind – in front) for all seven observers. The mean difference for the group is 52 ± 92 ms, which is significantly smaller than reaction time difference without separation ($1,316 \pm 474$ ms). This control experiment strongly supports our view that the difficulty arising from the L-shape-behind case is caused by surface completion behind the square, rendering target and distractor much less distinguishable at a surface representation level.

In these experiments, one concern is whether the stimuli used are appropriate to activate simple feature mechanisms differentially. There is no neurophysiological data to indicate the presence of 'filters' or units selectively tuned for L- and reversed L-shapes. However, simple orientation mechanisms would have different outputs for each pattern, which are likely to be sufficient for activation of different populations of orientation units. For example, Ls and Xs in many search tasks are easily distinguishable and a number of studies have indicated that a combination of low-level visual processing mechanisms could easily distinguish between these simple shapes^{12–14}. This assumption is supported by our third experiment, where we used a behavioural criterion which has been used to define a simple feature search, namely the existence of unvarying reaction times as distractor number is increased.

We used the same elements as the first experiment (Fig. 1a, b), varying the distractor number to see how this influenced reaction times in the two depth conditions. The relation between mean reaction times and number of distractors for two observers is shown in Fig. 3. There is a distinct qualitative difference in performance between the different depth conditions. When the L-shape is behind, reaction times increase dramatically as distractor number is increased. When the L-shape was seen in front, the slope is essentially flat with little or no increase in reaction time with increasing distractor number. Such flat functions are believed to indicate that search depends on discerning differences based on simple feature representations¹⁵.

Our results indicate that binocular disparity plays a large part in the slowing down of visual search but only in cases where it facilitates surface completion behind an adjoining occluder. When completion occurs, the target and distractor become perceptually more alike, in this case each appearing as a square, and thus become harder to distinguish than an L and a mirror-reversed L. This indicates that visual search is applied at a much higher level of visual representation than feature detection, most probably at the level of perceived shapes or surfaces. The result, that subtle changes which did not alter the features themselves had a profound effect in making visual search much more difficult, supports our original supposition that features cannot be accessed directly. The results also provide a possible explanation of the recent experiments showing that different polyhedral⁴ and shaded spherical forms⁵ are much easier to find in visual search experiments. Instead of assuming that such forms are

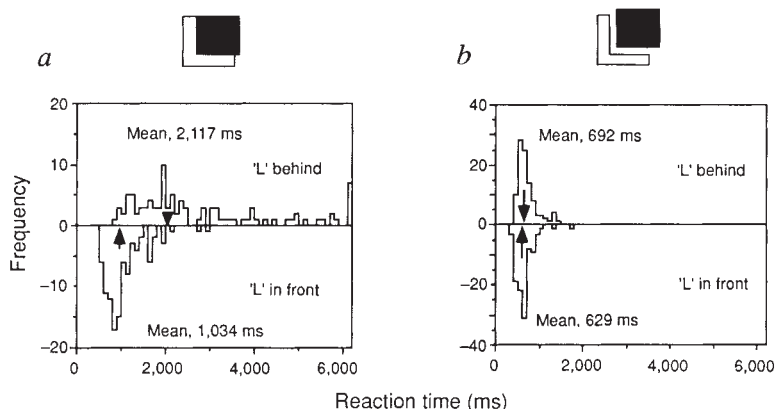


FIG. 2 Reaction time histograms (100 trials) of observer P.V. in the first experiment (a) where amodal completion of the L behind the square was possible and the second experiment (b) where it was not. The histograms above and below the horizontal axis represent reaction times obtained when Ls were behind and in front of the squares, respectively (disparity, 12.4'). Arrows indicate the median reaction times. In both experiments, 20 distractors were used.

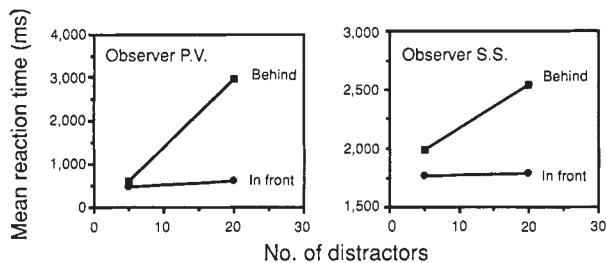


FIG. 3 Relation between mean reaction time (100 trials per point) and the number of distractors, with Ls placed either in front or behind.

simple features, we think it more appropriate to regard them as perceived three-dimensional shapes, distinguished at an object or surface level.

We attach particular significance to the fact that visual search is impaired rather than improved by changing a surface representation. It could be argued that the improvement in performance with particular three-dimensional shapes^{4,5} is not surprising, because information more meaningfully configured for

higher-level vision is also available to be searched. Here we draw an additional conclusion, that the feature-detection level of representation cannot be directly accessed by visual search mechanisms, and that in order to perform a visual search task, the system must operate at a higher level. □

Received 13 April; accepted 27 July 1992.

1. Julesz, B. *Nature* **290**, 91-97 (1981).
2. Treisman, A. & Gelade, G. *Cognitive Psychol.* **12**, 97-136 (1980).
3. Nakayama, K. & Silverman, G. H. *Nature* **320**, 264-265 (1986).
4. Enns, J. T. & Rensink, R. *Psychol. Sci.* **1**, 323-326 (1990).
5. Kleffner, D. & Ramachandran, V. *Percept. Psychophys.* **52**, 18-36 (1992).
6. Nakayama, K., Shimojo, S. & Silverman, G. H. *Perception* **18**, 55-68 (1989).
7. Nakayama, K. & Shimojo, S. *Cold Spring Harb. Symp. quant. Biol.* **40**, 911-924 (1990).
8. Kanisza, G. *Organization in Vision, Essays in Gestalt Perception* (Prager, New York, 1979).
9. Shimojo, S., Silverman, G. H. & Nakayama, K. *Vision Res.* **29**, 619-626 (1989).
10. Shimojo, S. & Nakayama, K. *Perception* **19**, 285-299 (1990).
11. Fox, R. & Patterson, R. *Percept. Psychophys.* **30**, 513-520 (1981).
12. Beck, J. *Percept. Psychophys.* **1**, 300-302 (1966).
13. Bergen, J. R. & Julesz, B. *Nature* **303**, 696-698 (1983).
14. Malik, J. & Perona, P. *J. opt. Soc. Am. A7*, 923-932 (1990).
15. Treisman, A. & Souther, J. *J. exp. Psychol. Gen.* **114**, 285-310 (1985).

ACKNOWLEDGEMENTS. This work was supported in part by a grant from Air Force Office of Scientific Research.

Molecular identification of the gene responsible for congenital nephrogenic diabetes insipidus

Walter Rosenthal*, Anita Seibold*, Anaid Antaramian*, Michèle Lonergan†, Marie-Francoise Arthus†, Geoffrey N. Hendy†, Mariel Birnbaumer*‡ & Daniel G. Bichet†

* Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030, USA

† Service de Néphrologie, Centre de Recherche, Hôpital du Sacre-Coeur de Montréal et Département de Médecine, Université de Montréal, Montréal, Québec H4J 1C5, Canada

ANTIDIURETIC hormone (arginine vasopressin) binds to and activates V2 receptors in renal collecting tubule cells. Subsequent stimulation of the G_s/adenylyl cyclase system promotes insertion of water pores into the luminal membrane and thereby reabsorption of fluid. In congenital nephrogenic diabetes insipidus (CNDI)¹, an X-linked recessive disorder, the kidney fails to respond to arginine vasopressin. Here we report that an affected male of a family with CNDI^{2,3} has a deletion in the open reading frame of the V2 receptor gene, causing a frame shift and premature termination of translation in the third intracellular loop of the receptor protein. A normal receptor gene was found in the patient's brother. Both the normal and the mutant allele were detected in his mother. A different mutation, causing a codon change in the third transmembrane domain of the V2 receptor, was found in the open reading frame of an affected male but not in the unaffected brother belonging to another family suffering from CNDI.

In one family (Q-5)^{2,3}, nephrogenic diabetes insipidus shows the typical X-chromosomal recessive inheritance (Fig. 1). The 37-year-old patient III₁ (of 32 years in ref. 2) has a history of polyuria and polydipsia, documented shortly after birth. During the first year of his life, he suffered repeated and prolonged episodes of dehydration with consequent mental retardation. Clinical investigations revealed the absence of renal and extrarenal V2 receptor-mediated responses in him and other affected members of the Q-5 family².

The open reading frame (ORF) plus 2 introns of the V2 receptor gene (2,025 base pairs (bp))⁴ was amplified by poly-

merase chain reaction (PCR)⁵ from genomic DNA of members of the Q-5 family and from healthy volunteers. Restriction enzyme analysis of PCR products from patient III₁ revealed no abnormalities (data not shown). PCR products of patient III₁ were cloned⁶ and sequenced⁷. Two deviations were found, both confined to the ORF. One was a transversion (A → G) at position 927 of the complementary DNA sequence⁸, which did not result in a codon change (not shown). The second was a deletion of one out of six consecutive guanosines (nucleotides 733-738) in the cDNA sequence⁸ (Fig. 2a). This deletion changes codon 246 from GGG to GGC, shifts the reading frame with an altered amino-acid sequence beginning with codon 247, and introduces a premature stop codon (TGA) at position 270 (Fig. 3). Thus, the mutant receptor lacks the entire carboxy-terminal third and conserves only 17 of the 42 amino acids of the third intracellular loop. An *in vitro* truncated human β 2-adrenoceptor, which like the V2 receptor couples to the G_s/adenylyl cyclase system, is inactive⁹. This strongly suggests that the Q-5 mutant is unable to transduce a signal. To our knowledge, this is the first example of a genetic defect in a G-protein-coupled hormone receptor. Genetic defects have been found in G-protein-coupled photo-receptor rhodopsin. Individuals with such defects are affected by autosomal dominant retinitis pigmentosa.¹⁰

Genetic linkage¹¹ and functional studies¹² have located the CNDI gene to the q28-qter portion of the X chromosome. To test whether the occurrence of the mutant gene is consistent with an X-linked mode of inheritance, we analysed the genes of the patient's healthy brother (III₂) and his mother (II₃).

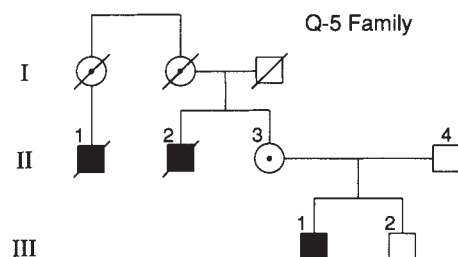


FIG. 1 Pedigree of the Q-5 family. Solid squares designate affected patients. Circles with a central dot are obligatory carriers. All affected members of the Q-5 family showed the same phenotype, that is no renal or extrarenal responses to arginine vasopressin or 1-desamino(8-D-arginine) vasopressin^{2,3}; female carriers were asymptomatic (D.G.B., unpublished results).

‡ To whom correspondence should be addressed.

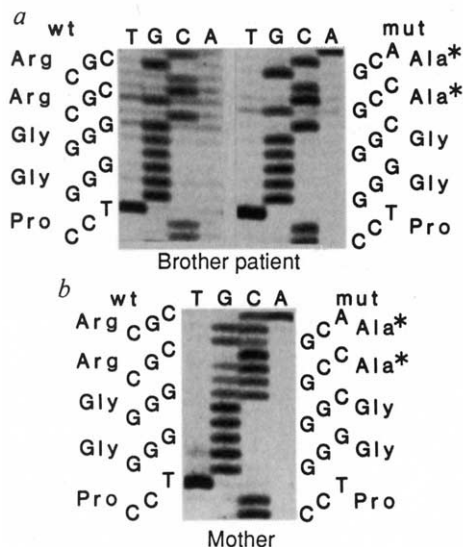
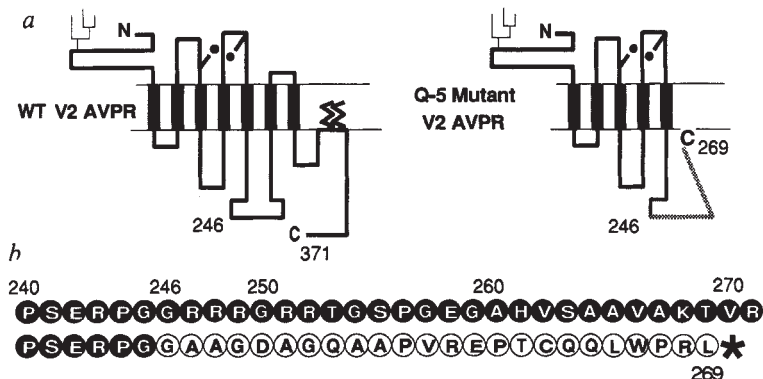


FIG. 2 Nucleotide sequences of the V2 receptor gene in members of the Q-5 family. *a*, Wild-type (wt) sequence of unaffected male sibling III₂ and mutant (mut) sequence of patient III₁. The V2 receptor gene was amplified by PCR from genomic DNA, cloned into the plasmid Bluescript and sequenced. Shown is the region containing the site of deletion. *b*, Sequence of the patient's mother (III₃). The portion of the V2 receptor gene containing the site of deletion was amplified by asymmetric PCR from genomic DNA of the mother and directly sequenced. Note that owing to heterozygosity, a single sequence is read up to the point of deletion and two sequences are read thereafter.

METHODS. Genomic DNA was prepared from white blood cells of healthy volunteers or members of the Q-5 family using standard techniques¹⁸. PCR primers flanking the coding region of the V2 receptor gene had the following sequences: 5'-CCAGGACTGGCCATACTG-3' (sense primer, corresponding to nucleotides -177 to -160) and 5'-CCAGCTCAGTGAGCTGAC-3' (antisense primer, corresponding to nucleotides +1,812 to +1,829). PCR was done with 0.5–1 µg genomic DNA serving as template in a mixture (100 µl)

Neither the transversion (not shown) nor the deletion (Fig. 2*a*) were detected in the brother, demonstrating that the patient and his healthy brother have inherited different alleles of the V2 receptor gene. Of 12 clones of the mother, seven showed the transversion and the deletion and five were normal (data not shown). We also amplified portions of the V2 receptor gene by asymmetric PCR¹³. Direct sequencing of PCR products from the patient or his healthy brother revealed either the mutant or the normal sequence, respectively (not shown), whereas sequencing of the PCR product of the mother (Fig. 2*b*), showed the sequences of both the wild-type and the mutant gene. These findings identify the asymptomatic mother as carrier of the mutant gene found in patient III₁.

FIG. 3 Models (*a*) and amino-acid sequences (*b*) of the wild-type and Q-5 mutant V2 receptor. *a*, Amino-acid sequences of the wild-type V2 receptor (WT V2 AVPR) and the Q-5 mutant V2 receptor (Q-5 mutant V2 AVPR) are shown as black and grey lines, respectively; putative transmembrane α -helices are shown as thick bars. On the basis of the presence of a consensus region for glycosylation in the predicted extracellular region of the receptor, the models depict an oligosaccharide residue at asparagine 22. A disulphide-bridge is postulated between the first and second extracellular loops based on mutational studies of Khorana *et al.* with rhodopsin²² and the conservation of cognate cysteines in all G-protein-coupled receptors. Folding of N and C termini and of longer intracellular loops is arbitrary. Anchoring of the carboxy terminal portion is patterned after those of rhodopsin (two palmitoyl anchors). *b*, Amino acids are given in one-letter code. The deletion in the Q-5 mutant causes a frame shift beginning with codon 246 (transition from filled to open circles; see text). The asterisk indicates a stop codon.



containing each primer at a concentration of 1.0 µM, 0.2 mM dNTPs, 2.5 units *Taq* polymerase (Perkin-Elmer Cetus), 1.5 mM MgCl₂, 50 mM KCl, 0.01% gelatin and 10 mM Tris-HCl (pH 8.3). After initial denaturation (3 min at 95°C), the following cycle was repeated 48 times: 45 s at 95°C, 2 min at 60°C (annealing) and 3 min with 6 s addition at each cycle at 72°C (extension); final extension was for 7 min at 72°C. PCR products were blunt-ended and cloned into *Sma*I-cut Bluescript KS (Stratagene) as described by Perez-Reyes *et al.*⁶. The insert:vector ratio in the ligation reactions varied from 0.5 to 5. The *Escherichia coli* strain DH1αF⁺ was transformed by using a modified version of Hanahan's high efficiency procedure¹⁹ provided by X. Wei, Baylor College of Medicine. Plasmids were isolated from bacterial cultures by alkaline lysis²⁰ and purified by polyethylene glycol precipitation⁶. Digests of plasmids with *Xba*I, *Sph*I and *Hind*III were used to confirm the presence of inserts. Sequencing of double-stranded DNA using the dideoxy chain-termination method⁷ was done as described⁶. Deviations from the reported sequence were confirmed by cloning and sequencing at least two independently obtained PCR products. The deletion found in the gene of patient III₁ was detectable with several of each sense and antisense primers used for sequencing. In the experiments shown in Fig. 2*a*, dITP replaced dGTP in the sequencing reaction; the sequence of the sense primer used was 5'-CTGATGGTGTTCGTGGCA-3' (corresponding to nucleotides 992–1,009 in the gene⁴). Asymmetric PCR¹³ was done as described by Gibbs *et al.*²¹. In the first reaction, 0.5 µg genomic DNA, 2 units of *Taq* polymerase and primers flanking the site of the deletion (1.0 µM each) were used. The sequences of the primers were: 5'-AACGTGGAAGGTGG-CAGC-3' (sense primer, corresponding to nucleotides +905 to +992 in the gene) and 5'-AGCCCTAGCCACGGCTACA-3' (antisense primer, nucleotides +1,277 to +1,295 in the gene). The assay volume was 50 µl; concentrations of other assay constituents were as described above. After initial denaturation (7 min at 95°C), the following cycle was repeated 25 times: 40 s at 96°C, 30 s at 60°C, 60 s at 72°C; final extension was for 7 min at 72°C. In the second reaction, 2 µl of the first PCR mixture were used as template; sense or antisense oligonucleotides were used as PCR primers in the second reaction. The sequence of the antisense primer used in the experiment shown (Fig. 2*b*) was 5'-AGCACAGCACATAGACGACCA-3' (nucleotides +1,191 to +1,211 in the gene). Compared to the first reaction, annealing and extension times were doubled; other assay conditions were not changed. Before sequencing, the products were extracted with chloroform and precipitated twice with ammonium acetate and ethanol. Sequencing of single-stranded DNA was done as described above.

Recently we located the V2 receptor gene to q28-qter portion of the X chromosome⁴. This and the discovery of the Q-5 mutant leads us to conclude that the CNDI gene is identical to the V2 receptor gene. It remains to be seen whether the deletion giving rise to the Q-5 mutant also occurs in other Canadian families with CNDI.

A CNDI patient from an unrelated family (O-1)^{2,3} was analysed as described above and found not to have the Q-5 mutation. But sequencing the complete ORF after cloning or sequencing of short products obtained by asymmetric PCR revealed a C → A transversion at nucleotide 395 of the cDNA sequence⁸. This point mutation changes codon 132 (third transmembrane domain) from GCC (alanine) to GAC (aspartic acid) and causes

a change in the hydropathy profile¹⁴ of the receptor protein. The defect was not found in the gene of the patient's brother, and both the normal and the mutant allele were amplified from the genomic DNA of the mother.

Patients with CNDI respond to adrenaline and parathyroid hormone (but not to arginine vasopressin or its analogues) with an increase in plasma cyclic AMP concentration and urinary cAMP excretion, respectively^{15,16}. Thus, genetic defects in ubiquitous signal transduction components (the G_s/adenylyl cyclase system) as they occur in pseudohypoparathyroidism type Ia¹⁷, are not likely. On the basis of our findings and the observation that CNDI patients generally lack both V2 receptor-mediated renal and extrarenal responses to arginine vasopressin², we predict that the disorder will frequently be ascribable to a defect of the V2 receptor gene. □

Received 9 June; accepted 3 August 1992.

- Reeves, W. B. & Andreoli, T. E. in *The Metabolic Basis of Inherited Disease* 6th edn (eds Scriver, C. R., Beaudet, A. L., Sly, W. S. & Valle, D.) 1985–2011 (McGraw-Hill, New York, 1989).
- Bichet, D. G. *et al.* *New Engl. J. Med.* **318**, 881–887 (1988).

- Bichet, D. G. *et al.* *Am. J. hum. Genet.* (in the press).
- Seibold, A., Brabet, P., Rosenthal, W. & Birnbaumer, M. *Am. J. hum. Genet.* (in the press).
- Saiki, R. K. *et al.* *Science* **239**, 487–491 (1988).
- Perez-Reyes, E., Wei, X., Castellano, A. & Birnbaumer, L. *J. biol. Chem.* **265**, 20430–20436 (1990).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Birnbaumer, M. *et al.* *Nature* **357**, 333–335 (1992).
- Koblick, B. K. *et al.* *Science* **240**, 1310–1316 (1988).
- Sung, C.-H., Schneider, B. G., Agarwal, N., Papermaster, D. S. & Nathans, J. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8840–8844 (1991).
- Knoers, N. *et al.* *Nephron* **50**, 187–190 (1988).
- Jans, D. A., van Oost, B. A., Ropers, H. H. & Fahrenholz, F. *J. biol. Chem.* **265**, 15379–15382 (1990).
- Gyllenstein, U. B. & Ehrlich, H. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7652–7656 (1988).
- Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105–132 (1982).
- Bichet, D. G. *et al.* *Kidney Int.* **36**, 859–866 (1989).
- Moses, A. M. & Coulson, B. B. *J. clin. Endocr. Metab.* **55**, 699–702 (1982).
- Patten, J. L. *et al.* *New Engl. J. Med.* **322**, 1412–1419 (1988).
- DiLella, A. G. & Woo, S. L. C. *Meth. Enzym.* **152**, 199–212 (1987).
- Hanahan, D. in *DNA Cloning* (ed. Glover, D. M.) 109–135 (IRL, Oxford, 1985).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning* (Cold Spring Harbor Laboratory Press, New York 1989).
- Gibbs, R. A., Nguyen, P.-N. & Caskey, C. T. in *Methods in Molecular Biology* Vol. 8 (ed. Mathew, C.) 9–20 (Human Press, Clifton, New York, 1991).
- Karnik, S. S., Sakmar, T. P., Chen, H.-B. & Khorana, H. G. *Biochemistry* **85**, 8459–8463 (1988).

ACKNOWLEDGEMENTS. We thank L. Birnbaumer, P. Brabet, J. Codina, R. Coria and U. Rudolph for support and helpful discussions, and A. Didwania for help in sequencing the O-1 family. This work was supported in part by NIH grants to M.B. and a grant from the Medical Research Council of Canada to D.G.B. W.R. is a recipient of a Heisenberg Fellowship from the Deutsche Forschungsgemeinschaft.

APC mutations occur early during colorectal tumorigenesis

Steven M. Powell*, Nathan Zilz*,
Yasmin Beazer-Barclay*, Tracy M. Bryan*,
Stanley R. Hamilton†, Stephen N. Thibodeau‡,
Bert Vogelstein* & Kenneth W. Kinzler*§

* The Johns Hopkins Oncology Center, 424 North Bond Street, Baltimore, Maryland 21231-1001, USA

† Department of Pathology, Johns Hopkins School of Medicine, 720 Rutland Avenue, Baltimore, Maryland 21205-2196, USA

‡ Mayo Clinic, Department of Laboratory Medicine, Rochester, Minnesota 55905, USA

HUMAN tumorigenesis is associated with the accumulation of mutations both in oncogenes and in tumour suppressor genes^{1–3}. But in no common adult cancer have the mutations that are critical in the early stages of the tumorigenic process been defined. We have attempted to determine if mutations of the APC gene play such a role in human colorectal tumours, which evolve from small benign tumours (adenomas) to larger malignant tumours (carcinomas) over the course of several decades. Here we report that sequence analysis of 41 colorectal tumours revealed that the majority of colorectal carcinomas (60%) and adenomas (63%) contained a mutated APC gene. Furthermore, the APC gene met two criteria of importance for tumour initiation. First, mutations of this gene were found in the earliest tumours that could be analysed, including adenomas as small as 0.5 cm in diameter. Second, the frequency of such mutations remained constant as tumours progressed from benign to malignant stages. These data provide strong evidence that mutations of the APC gene play a major role in the early development of colorectal neoplasms.

The candidate tumour suppressor gene, APC^{4,5}, is abnormal in the germline of patients with familial adenomatous polyposis (FAP), an inherited disease predisposing to colorectal neoplasms^{6–8}. But most colorectal neoplasms occur in patients without FAP. Until now, the role of the APC protein (APC) in these more common neoplasms has been largely conjectural because only a few APC mutations, occurring in advanced colorectal carcinomas, have been described⁶. To determine the role of APC in the development of sporadic colorectal tumours, we examined the APC gene for mutations in 41 sporadic

TABLE 1 Somatic APC mutations

Sample*	Size or stage†	Alleles‡	Nucleotide change§	Coding change
Ca-1	B	2	CGA → TGA	Arg → Stop (Codon 564)
Ca-2	C	2	taatttttag/GGT → tagtttttag/GGT	Splice acceptor (Exon 8)
Ca-4	A	2	AAA → AA	Deletion (Codon 2,455)
Ca-6	B	2	CGA → TGA	Arg → Stop (Codon 876)
Ca-8	B	2	CCT → CC	Deletion (Codon 1,443)
Ca-10	B	2	AAA → AA	Deletion (Codon 758)
Ca-12	B	2	TTT → TT	Deletion (Codon 801)
Ca-14	B	2	AAT → AT	Deletion (Codon 1,455)
Ca-16	B	2	AAG → G	Deletion (Codon 1,462)
Ca-18	B	2	CGA → TGA	Arg → Stop (Codon 232)
Ca-19	C	1	CGA → TGA	Arg → Stop (Codon 564)
Ca-20	C	2	CGA → TGA	Arg → Stop (Codon 1,450)
Ca-21	B	2	AAG → G	Deletion (Codon 1,462)
Ca-22	B	2	AAT → AT	Deletion (Codon 869)
Ca-25	C	1	TTT → T	Deletion (Codon 773)
Ad-1	1.1 cm	2	ATA → AATA	Insertion (Codon 1,307)
Ad-2	0.8 cm	2	17 bp deletion (nt 4060–4077)	Deletion (Codon 1,354)
Ad-3	0.7 cm	1	GTT → TT	Deletion (Codon 1,405)
Ad-4	1.2 cm	2	AAA → TAA	Cys → Stop (Codon 670)
Ad-7	2.5 cm	2	TCT → TAT	Ser → Tyr (Codon 906)
Ad-9	0.5 cm	2	CGA → TGA	Arg → Stop (Codon 1,450)
Ad-10	1.2 cm	2	TCA → TAA	Ser → Stop (Codon 1,346)
Ad-11	0.7 cm	1	CGA → TGA	Arg → Stop (Codon 876)
Ad-12	0.5 cm	2	ACT → AACT	Insertion (Codon 1,556)
Ad-15	3.5 cm	1	GAA → AA	Deletion (Codon 403)
			AGT → AGAGAGT	Insertion (Codon 1,465)
			tag/GTC → tag/GTC	Splice acceptor (Exon 4)
			CGA → TGA	Arg → Stop (Codon 405)
			ACT → AACT	Insertion (Codon 1,556)
			CAG → TAG	Gln → Stop (Codon 1,367)
			CCA → ACCA	Insertion (Codon 1,941)
			CCT → CCTT	Insertion (Codon 1,373)
			TCA → TAA	Ser → Stop (Codon 1,281)
			CGA → TGA	Arg → Stop (Codon 1,450)
			GAA → TAA	Glu → Stop (Codon 1,408)

* Ad indicates an adenoma and Ca indicates a carcinoma.

† For carcinomas, the modified Dukes' stage¹⁷ is indicated. For adenomas, the largest diameter is given.

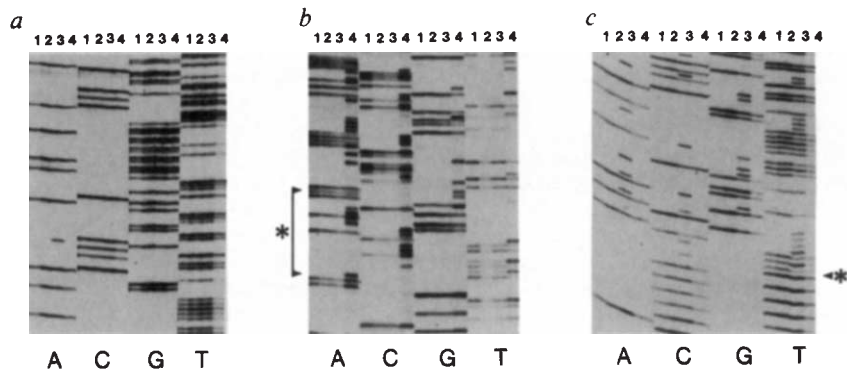
‡ Allelic statuses were determined using two dinucleotide repeat polymorphisms located on 5q^{18,19}, a deletion/insertion polymorphism 3' to exon 10 of MCC^{20,21}, and the APC sequence polymorphisms detailed in Table 2.

§ Underlined nucleotides were mutant. Lower case letters represent introns, upper case letters represent exons.

|| The first affected codon or exon is listed in parentheses.

FIG. 1 Screening for *APC* mutations in sporadic colorectal tumours by DNA sequencing. Three representative examples of mutations are illustrated. Autoradiographs of sequencing gels from four different tumours (labelled 1 to 4) are shown in each panel. Sequencing reactions were grouped by nucleotide (A, C, G and T) to aid identification of changes. Carcinoma Ca-22 (lane 3 in *a*) contained a C→T transition at the first base of codon 1,450 creating an 'A' (marked with an asterisk) in the antisense sequence displayed. Carcinoma Ca-19 (lane 4 in *b*) contained a 17 bp deletion affecting codons 1,354 to 1,359 as indicated by an asterisk and bracket on the sense sequence displayed. Adenoma Ad-3 (lane 3 in *c*) contained a 4 bp insertion beginning at codon 1,465 creating a frameshift beginning at the base marked with an asterisk in the antisense sequence shown.

METHODS. DNA was prepared from tumours enriched for neoplastic cells by cryostat sectioning as previously described²¹. Normal DNA was prepared from adjacent nonaffected colonic mucosa. Exons 1 through 14 and surrounding splice sites were individually amplified using PCR and subcloned into a T-tailed pBluescript as previously described⁶. For exon 15, three separate but overlapping fragments, each ~2.3 kb, were amplified with



paired primers engineered to contain either an *EcoRI* or *BamHI* sites at their ends to aid directional subcloning into a *BamHI*- and *EcoRI*-digested pBlue-script vector. Pools of at least 50 clones for each PCR product were sequenced with internal primers as previously described²². The 34 amplification primers and 49 sequencing primers were designed from previously reported sequence data^{4,7,8} and their sequences are available from the authors on request. Each mutation was confirmed by a second independent PCR amplification and sequence analysis.

colorectal tumours of various size, stage and degree of malignancy. These tumours included 25 carcinomas ranging from localized disease to fully metastatic disease and 16 adenomas ranging from 0.5 cm to 4.5 cm (average 1.6 cm) in diameter. Although rapid screening techniques are often used to search for mutations in genes as large as *APC*, we chose to sequence the entire coding region of the gene in all 41 tumours. Although more difficult, the sequencing approach provides maximum sensitivity for detecting coding and splice-site mutations. For this purpose, individual exons and their surrounding splice sites were amplified from tumour DNA using the polymerase chain reaction (PCR). The amplified DNA was then cloned and sequenced using internal sequencing primers (see legend to Fig. 1). In each case wherein a variant was identified, an independent amplification was used to confirm its validity, and its definition as a somatic mutation was accomplished by comparing the sequence to that found in normal tissues of the same patient.

Fifteen of the 25 (60%) colorectal carcinomas and 10 of 16 (63%) adenomas examined contained at least one somatic mutation of the *APC* gene (Table 1, examples in Fig. 1). Most of the mutations resulted in a truncated *APC* protein. There was similarity in the character and distribution of mutations in both the adenomas and carcinomas (Table 1). Of the 35 mutations identified, six were insertions ranging from 1 to 4 base pairs (bp) and 12 were deletions ranging from 1 to 17 bp. All of the insertions and deletions produced a frameshift that would truncate the protein. Similarly, 14 of the point mutations found were nonsense mutations whereas only one was a missense mutation. The two remaining somatic mutations altered splice sites and would be expected to result in an *APC* transcript containing a frameshift. One abolished the splice acceptor site for exon 4 (Ad-4, Table 1) and the other created a cryptic splice acceptor sequence 7 bp upstream of exon 8 (Ca-2, Table 1). Of the 17 point mutations, 15 (88%) were at G-C bp and 9 of these were at a CpG dinucleotide. This same tendency for mutations to occur at G-C bp and particularly at CpG dinucleotides was also observed in the *p53* gene from colorectal tumours⁹, suggesting deamination of 5-methylcytosine as a possible mechanism¹⁰.

The majority (90%) of the mutations identified were distributed over the first 55% of the coding region, with half of the mutations occurring in a 722 bp region spanning codons 1281 to 1554 (Fig. 2). These somatic mutations in sporadic colorectal tumours are very similar to those found in the germ line of FAP patients in terms of both type (tuncations) and distribution along the gene⁸.

Many of the tumours had alterations of both alleles of their *APC* gene. In five tumours, one copy of the *APC* gene was inactivated by allelic loss and the remaining copy contained an intragenic mutation (Ca-19, Ca-25, Ad-3, Ad-11, Ad-15, Table 1). An additional nine tumours contained two intragenic mutations, presumably one on each allele (Table 1). This presumption was testable in three of these nine cases because the mutations were close enough to be included in one PCR product (Ca-10, Ca-14, Ca-18, Table 1). In all three cases, sequence analysis of individual PCR clones indicated that the mutations were on separate alleles. One tumour had three somatic mutations (Ad-4, Table 1). Histopathology of this tumour revealed two morphologically distinct areas of adenomatous growth suggesting that two independent areas of clonal expansion were present in the sample analysed. In total, 14 of 41 tumours (31% of adenomas, 36% of carcinomas) were predicted to have no normally functioning *APC* protein. This suggests that there is a growth advantage afforded by mutating both copies of the *APC* gene but that two mutations may not be required for tumorigenesis.

The fact that colorectal adenomas (63%) were just as likely as colorectal carcinomas (60%) to acquire an *APC* mutation suggests that *APC* mutations occur as an early event during colorectal tumorigenesis. Indeed, *APC* mutations were detected in the smallest adenomas that were examined, including five tumours less than 1 cm in diameter (Table 1). We have examined these same five tumours for mutations in the three *ras* genes (*K-ras*, *N-ras*, and *H-ras*), which were heretofore the only genes known to be specifically mutated during the early stages of colorectal tumorigenesis¹¹. Only one of the five tumours with *APC* gene mutations contained a *ras* gene mutation, consistent with the idea that *APC* alterations precede those of *ras*.

FIG. 2 Distribution of *APC* mutations. This diagram illustrates the position of all 35 somatic mutations relative to the *APC* protein and exon structure. The boxes correspond to exons (1, 3, 5, 7, 9, 11, 14 and 15 are numbered for reference) and the numbers below represent amino-acid residues. Truncations arose from nonsense mutations, insertions or deletions as detailed in Table 1.

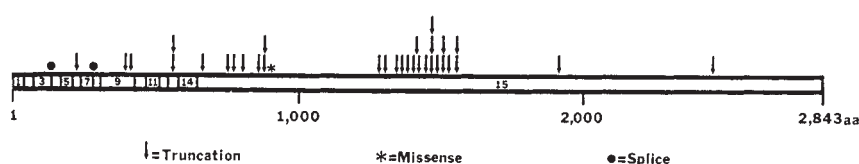


TABLE 2 Germ line *APC* variants

Frequency*	Nucleotide change†	Coding change‡
57%/43%	TAC/TAT	Silent (Codon 486)§
62%/38%	GCA/GCG	Silent (Codon 545)§
63%/37%	ACA/ACG	Silent (codon 1,493)§
63%/37%	TCG/TCI	Silent (Codon 1,756)
63%/37%	CCA/CCG	Silent (Codon 1,960)
66%/34%	GGA/GGG	Silent (Codon 1,678)§
96%/4%	GGA/GGG	Silent (Codon 2,568)
90%/10%	GAC/GTC	Asp/Val (Codon 1,822)
98%/2%	CCA/TCA	Pro/Ser (Codon 870)
99%/1%	GGA/GTA	Gly/Val (Codon 84)

* Determined by sequence analysis of the 90 alleles in 45 individuals.

† Underlined nucleotides were variant.

‡ The codon affected is listed in parentheses.

§ Reported previously in ref. 8.

|| Reported previously in ref. 7.

During the course of these studies, several germline variants were identified, most of which were common silent polymorphisms (Table 2). But three variants of *APC* were observed that resulted in amino-acid changes. One variant produced an aspartate to valine change at codon 1,822 and was found in nine of the patients examined. A second variant was a proline to serine change at codon 870, found in two patients. The third was a glycine to valine change at codon 84, found in a single patient. Analysis of the allele frequencies of these three variants using allele specific hybridization in 100 healthy individuals revealed no significant difference between the normal population and the patient population examined initially (data not shown). This is interesting in light of the suggestion that there is a common inherited predisposition to colorectal neoplasia affecting 19% of the population¹². At this point, it is not possible to determine whether these amino-acid changes result in an inherited colorectal tumour risk or are simply inconsequential variants.

Our studies probably underestimate the true prevalence of *APC* abnormalities in sporadic colorectal tumours for two reasons. First, the current analysis did not include the promoter region or intronic portions of the *APC* gene. Mutations localized in these regions could also result in loss of *APC* activity. Second, this analysis only considered intragenic mutations of the *APC* gene, not large deletions that could also inactivate the *APC* gene. Indeed, allelic loss analysis revealed three additional tumours that had lost one copy of the *APC* gene but did not contain a detectable *APC* mutation; bringing the total to 28 (68%) tumours that had either mutated or lost at least one copy of the *APC* gene.

The high prevalence of *APC* gene mutations in the colorectal tumours examined provides strong *in vivo* evidence that *APC* alterations play a major role in the development of most common colorectal tumours. Furthermore, the data show that *APC* mutations are a very early event in colorectal tumorigenesis, if not the initiating event. The similar frequency of *APC* mutations as tumours progress from benign to malignant stages stands in marked contrast to other changes, such as inactivation of the p53 tumour suppressor gene, which increases in frequency as tumours advance¹³. These data also indicate that the small adenomas occurring in FAP patients and those that occur sporadically have a similar molecular defect: a mutation in *APC*, inherited in the former case and somatically acquired in the latter. This similarity has important practical as well as biological implications. First, nonsteroidal anti-inflammatory drugs reverse the growth of colorectal adenomas in FAP patients¹⁴. Such drugs might also find use in chemotherapy of sporadic tumours if the underlying pathogenesis is the same. Second, an animal model for FAP has been described in the form of a mouse lineage which contains a germ line mutation of the *APC* gene¹⁵. Because

of this shared genetic defect, tumours in these mice may prove to be a good model for sporadic colorectal tumours as well as for FAP. Finally, a potential screening test for colorectal tumours based on the detection of mutated genes from tumour cells shed into the stool was recently described¹⁶. Such a test might optimally be done using probes for genes that are mutated frequently and early in the neoplastic process, long before malignancy. The *APC* gene might therefore provide special opportunities in this regard. □

Received 17 June; accepted 31 July 1992.

1. Fearon, E. R. & Vogelstein, B. *Cell* **61**, 759–767 (1990).
2. Weinberg, R. A. *Science* **254**, 1138–1146 (1991).
3. Bishop, J. M. *Cell* **64**, 235–248 (1991).
4. Kinzler, K. W. *et al. Science* **253**, 661–665 (1991).
5. Joslyn, G. *et al. Cell* **66**, 601–613 (1991).
6. Nishisho, I. *et al. Science* **253**, 665–669 (1991).
7. Groden, J. *et al. Cell* **66**, 589–600 (1991).
8. Miyoshi, Y. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 4452–4456 (1992).
9. Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. C. *Science* **253**, 49–53 (1991).
10. Rideout, W. M., Coetzee, G. A., Olumi, A. F. & Jones, P. A. *Science* **249**, 1288–1290 (1990).
11. Vogelstein, B. *et al. New Eng. J. Med.* **319**, 525–532 (1988).
12. Cannon-Albright, L. A., Skolnick, M. H., Bishop, T., Lee, R. G. & Burt, R. W. *New Engl. J. Med.* **319**, 533–537 (1988).
13. Baker, S. J. *et al. Cancer Res.* **50**, 7717–7722 (1990).
14. Waddell, W. R., Gansler, G. F., Cerise, E. J. & Loughry, R. W. *Am. J. Surg.* **156**, 175–179 (1989).
15. Su, L.-K. *et al. Science* **256**, 668–670 (1992).
16. Sidransky, D. *et al. Science* **256**, 102–105 (1992).
17. Turnbull, R. B., Kyle, K., Watson, F. R. & Spratt, J. *Ann. Surg.* **166**, 420–427 (1967).
18. Spirio, L., Joslyn, G., Nelson, L., Leppert, M. & White, R. *Nucleic Acids Res.* **19**, 6348 (1991).
19. Wijnen, J. *et al. Nucleic Acids Res.* **19**, 6965 (1991).
20. Kinzler, K. W. *et al. Science* **251**, 1366–1370 (1991).
21. Boynton, R. F. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 3385–3388 (1992).
22. Nigro, J. M. *et al. Nature* **342**, 705–708 (1989).

Evidence that Hensen's node is a site of retinoic acid synthesis

Brigid L. M. Hogan*, Christina Thaller† & Gregor Eichele‡

* Department of Cell Biology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232-2175, USA

† V. & M. McLean Department of Biochemistry, Baylor College of Medicine, Houston, Texas 77030, USA

HENSEN'S node of amniotes, like the Spemann organizer of amphibians, can induce a second body axis when grafted into a host embryo¹. The avian node, as well as several midline structures originating from it (notochord, floor plate), can also induce digit pattern duplications when grafted into the chick wing bud^{2,3}. We report here that the equivalent of Hensen's node from mouse is an effective inducer of digits in the chick wing bud. Tissues anterior and posterior to the node also evoke pattern duplications, but with a significantly lower efficiency. The finding that the murine node operates in an avian wing bud suggests that the same inducing agent(s) function in both primary and secondary embryonic fields and have been conserved during vertebrate evolution. Digit pattern duplications are also evoked by local administration of all-trans-retinoic acid^{4,5}. This similarity raises the possibility that Hensen's node is a source of retinoic acid. The mouse node is capable of synthesizing retinoic acid from its biosynthetic precursor all-trans-retinol at a substantially higher rate than either anterior or posterior tissues.

To map the digit-inducing capacity of mouse embryonic tissues, intact embryos or dissected fragments ranging from the pre-streak, egg cylinder stage (6.5 days) to embryos with 2–3 somites (7.75 days) were implanted into chick wing buds. Egg cylinder stage embryos failed to induce additional digits (data

‡ To whom correspondence should be addressed.

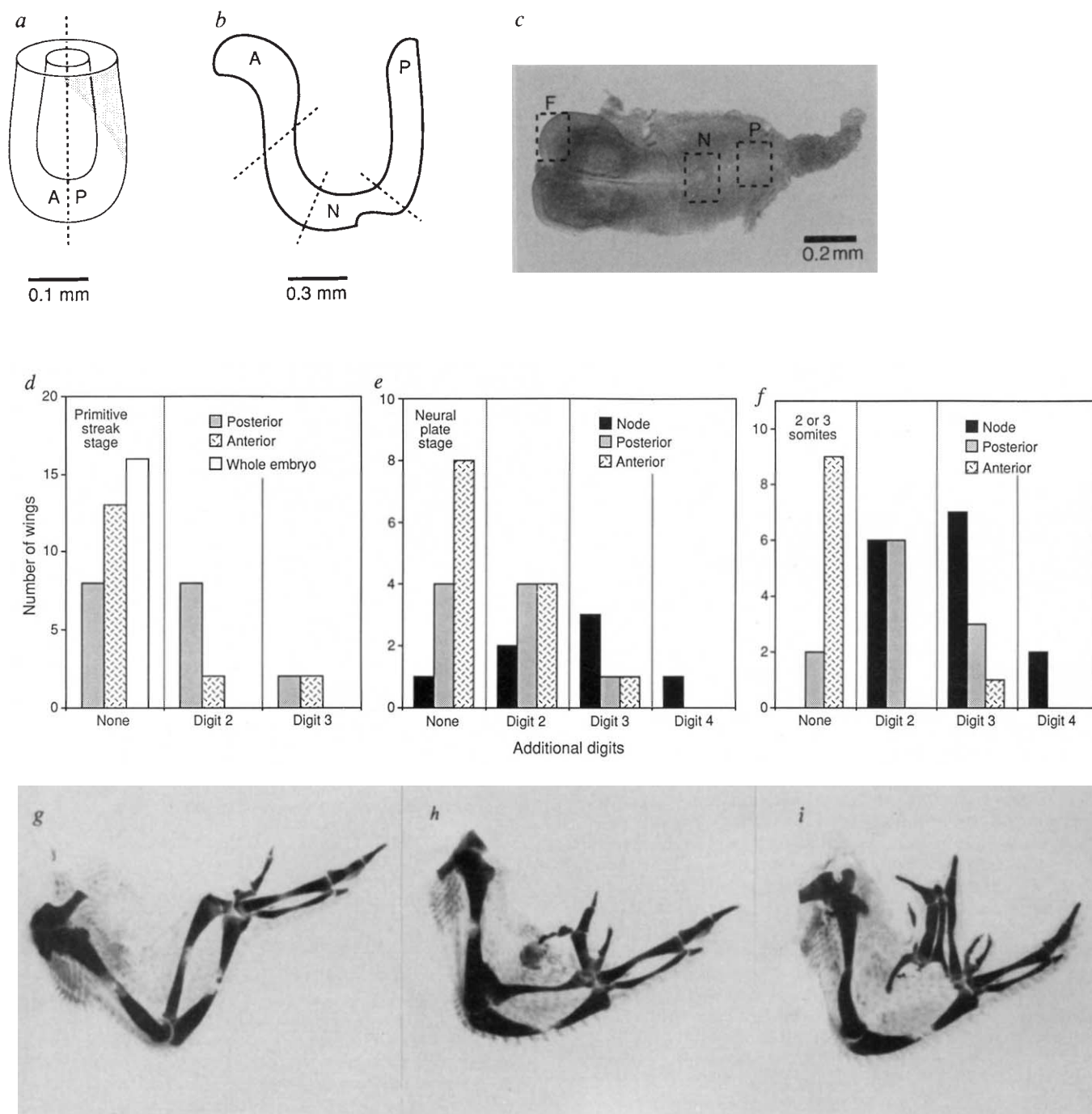


FIG. 1 Digit pattern duplications are induced by grafts of embryonic mouse tissues. *a*, Primitive streak stage embryo with a shield of mesoderm visible on the posterior side of the embryo (shaded). Dotted line represents a cut made to separate anterior (A) and posterior (P) halves. *b*, Schematic representation of a neural plate embryo (0–1 somites) viewed from the side. Four pieces of tissue were generated, fragments marked A (anterior), N (node) and P (posterior) were used for grafting. *c*, Photomicrograph of a 3-somite stage embryo indicating anterior (presumptive forebrain, F), node (N) and posterior (primitive streak, P) pieces used for grafting. *d–f*, Graphical representation of the digit duplication patterns evoked by primitive streak stage tissue (*d*), neural plate stage tissue (*e*) and 2–3-somite stage tissue (*f*). For generating histograms, only the highest-order digit ($4 > 3 > 2$) was taken into account. Thus the column labelled 'additional digit 3' represents for example 3234 or 32234 patterns. *g–i*, Digit patterns that developed after grafting tissues from a 2–3-somite stage embryo; *g*, normal 234 pattern, graft was presumptive forebrain neuroectoderm; *h*, 3234 pattern, graft was posterior tissue, arrow points at a piece of cartilage of unidentified nature; *i*, 23432234 pattern caused by a node graft. The specimen in *i* emphasizes that patterns obtained may lack mirror-symmetry usually observed

with grafts of zones of polarizing activity^{2,30}. *i* lacks a radius. No radii or deformed radii were found in about 60% of the wings, irrespective of the origin of the graft.

METHODS. Mouse embryos were dissected from ICR females mated with (C57BL × DBA) F₁ males between about 6.5 and 7.75 days p.c. (noon on the day of plug is 0.5 days p.c.) in Dulbecco's modified Eagle's medium (DMEM) with 10% calf serum and 25 mM HEPES buffer (pH 7.4). Embryos were separated from extraembryonic tissues, including the amnion and most of the yolk sac. Neural fold and somite stage embryos were flattened ventral side down on the surface of the dish and fragments isolated using tungsten needles. Cuts were placed to exclude the remaining extraembryonic tissues. Chick embryos used were at Hamburger–Hamilton stage 20. The anterior portion of the apical ectodermal ridge of the right wing bud was lifted (see ref. 3). Mouse tissue was picked up with a 0.3 mm diameter X-ray glass capillary and positioned underneath the apical ectodermal ridge. No attempts were made to orient the graft with respect to the ridge. The size of the graft was about 1/20 of that of the wing bud. After 7 days of incubation at 38 °C, embryos were fixed and cartilage was stained with Alcian Green.

not shown). Intact primitive streak embryos (Fig. 1a) were also negative (Fig. 1d). But dissecting these embryos into anterior and posterior halves and grafting them into host wing buds, produced patterns that have either an additional digit 2 or a 2 and a 3 (2234 or 32234) in 40% of the cases (Fig. 1d). Posterior tissue containing the primitive streak was slightly more efficient than anterior portions. Apparently, direct contact between the inside of the embryo and host tissue is advantageous for the inducer to be effective. When tissue from neural fold (Fig. 1b) or 2-3 somite embryos (Fig. 1c) was used, the percentage of wing patterns with additional digits increased to 52% and 70%, respectively. To determine whether inducing activity is localized in these embryos, they were dissected into an anterior piece, a fragment containing the node and a posterior part (Fig. 1b, c). Node tissue from neural plate and 2-3-somite embryos is a more effective inducer of pattern duplications than either anterior or posterior tissue (Fig. 1e, f). Out of 22 node grafts, 10 wings exhibited a 32234 or a 3234 pattern (Fig. 1h). Three wings had a pattern with additional digits 2, 3 and 4 (Fig. 1i); they were all derived from buds that had obtained node implants. In contrast, the response elicited by anterior or posterior tissues is weaker, most frequently resulting in 2234 patterns. The mouse node is no longer a distinct structure by the 8-somite stage. But in a small series of experiments with 4-8-somite embryos, slightly higher duplicating activity was seen with the presumed node region and posterior tissue than with anterior tissue (data not shown).

Studies in the chick wing have shown that retinoic acid can elicit the same types of pattern duplications as the node^{2,4-6}, raising the possibility that the node releases retinoic acid when implanted into the wing bud. If so, the node should be capable of generating retinoic acid from its precursor, all-*trans*-retinol^{7,8}. Moreover, weak inducers such as anterior neural fold should have less converting activity. To test this hypothesis, 10 to 17 equally sized pieces of anterior neural tissue, node or posterior tissue from 3-4-somite embryos (Fig. 1c for origin of fragments) were incubated in serum-free medium containing 500 nM [³H]retinol. After 5 h of incubation, extracts of the samples were prepared and fractionated over a C₁₈ reversed-phase high-performance liquid chromatography (HPLC) column. Figure 2a shows sample chromatograms for each tissue and for a control in which no tissue was added. These chromatograms

are very similar. The majority of the radioactivity remains in the form of all-*trans*-retinol, except for some polar metabolites (fractions 3-8) and minor peaks centred at fractions 17 and 37. Tissues from early mouse embryos do not generate 3,4-didehydroretinol, a major metabolite produced in chick limb buds incubated in retinol^{7,8}. Note that the node extract, but not the control, generates a small but distinct peak at the elution position of retinoic acid (insert Fig. 2a). To analyse this putative retinoic acid peak further, fractions 30 to 33 were pooled and rechromatographed on an identical C₁₈ column. The chromatogram of node extract (Fig. 2b) shows a distinct radioactivity peak co-eluting with authentic retinoic acid. A substantially smaller peak is obtained from posterior tissue (Fig. 2b). Anterior neuroectoderm, and the control sample lacking tissue, exhibit essentially background levels of radioactivity. The results of four independent conversion experiments are compiled in Table 1. The estimated amount of retinoic acid present in the node sample is 4 ± 0.85 pg per μ g DNA, which is 12 times more than in anterior tissue (Table 1). We conclude from these experiments that node and to a lesser extent posterior tissue from 3-4-somite mouse embryos are capable of converting all-*trans*-retinol to retinoic acid.

Several lines of evidence suggest that retinoic acid is a morphogenetic signalling molecule in vertebrate embryos and, in particular, plays a role in organizing pattern along the anteroposterior body axis. First, in mouse and chick, β -retinoic acid receptor is expressed in the node and in tissue posterior to it^{9,10}. In addition, transcripts for CRBP-I, a cellular protein that binds retinol, are present in the primitive streak region of the mouse embryo from at least day 7.5 *post coitum* (p.c.)⁹. CRBP-I might enrich retinol in and around the tissues capable of converting retinol to retinoic acid. Second, mouse embryos carrying a retinoic acid response element/*lacZ* reporter transgene do not express detectable levels of *lacZ* until around 7.5 days p.c., when digit duplication activity increases in our assays¹¹. Expression of *lacZ* is initiated along the midline, with an anterior limit at the level of the node. Third, exposing amphibian, avian and mammalian embryos to exogenous retinoic acid severely affects anteroposterior patterning¹²⁻¹⁹. In particular, when added during the time when expression patterns of homeobox-containing genes are established, retinoic acid alters these patterns in reproducible ways. For example, exposing 7.3-8-days p.c. mouse embryos to retinoic acid *in utero*,

TABLE 1 *In vitro* synthesis of all-*trans*-retinoic acid from all-*trans*-retinol by various embryonic tissues

Tissue type	Experiment	Number of pieces	All- <i>trans</i> -retinoic acid per piece (pg) [†]	DNA per piece (ng)	All- <i>trans</i> -retinoic acid per μ g DNA (pg)	Average and s.d. all- <i>trans</i> -retinoic acid per μ g DNA (pg)
Anterior	1	10	0	299	0	0.33 $\pm$ 0.39
	2	10	0	262	0	
	3	17	0.13	218	0.60	
	4*	15	0.20	273	0.73	
Node	1	10	0.94	287	3.28	3.96 $\pm$ 0.85
	2	10	0.86	264	3.26	
	3	17	1.20	240	5.00	
	4*	15	1.19	276	4.31	
Posterior	1	10	0.12	263	0.46	0.99 $\pm$ 0.49
	2	10	0.22	248	0.89	
	3	17	0.23	241	0.95	
	4*	15	0.43	261	1.65	

* Actual data shown in Fig. 2b.

[†] Retinoic acid (pg) per tissue piece is calculated as: $\Sigma[Ri - (i \times bgr)] / (SA \times m)$; where i is the number of fractions comprising the retinoic acid peak (3 in Fig. 2b); Ri is the amount of radioactivity per fraction (in d.p.m.); bgr is the average background radioactivity in 3 fractions following the retinoic acid peak; m is the number of embryo pieces used; SA is specific activity of retinol (287 d.p.m. μ g⁻¹).

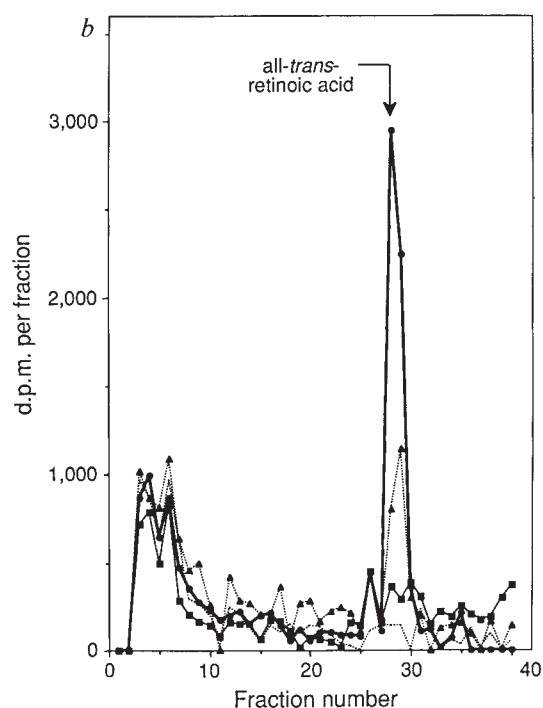
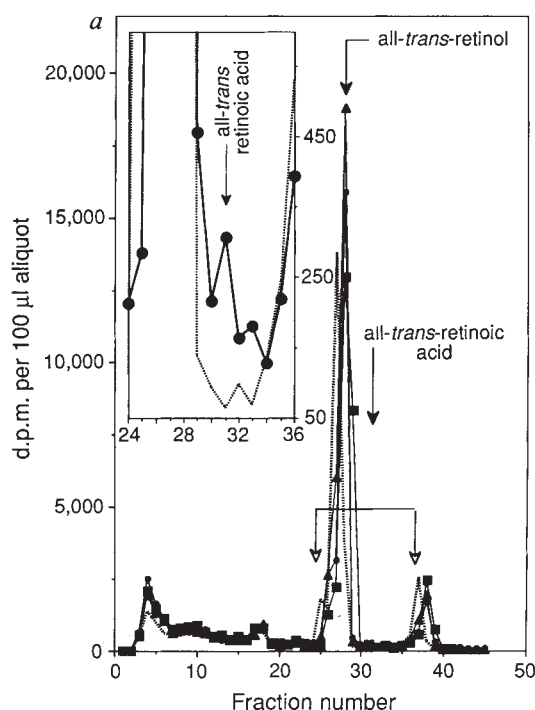
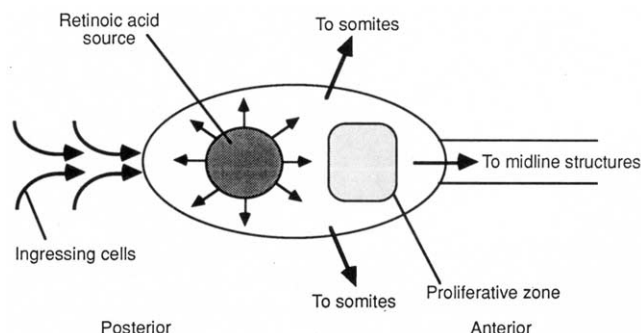


FIG. 2 Sample chromatograms illustrating the conversion of [^3H]-labelled all-*trans*-retinol to [^3H]-labelled all-*trans*-retinoic acid by particular tissue pieces from 3–4-somite embryos. *a*, The radioactivity profiles generated by anterior tissue (squares), posterior tissue (triangles), node (filled circles) and a control incubate without tissue (dotted line). The insert shows fractions 24 to 36 (marked by a bracket in the main figure) with an expanded ordinate. Node tissue but not the control generates a readily detectable peak coeluting with authentic retinoic acid. *b*, Refractionation of fractions 30 to 33 in *a*. In the case of the node sample (filled circles), a distinct peak appears that coelutes with internal standard retinoic acid. Posterior tissue (triangles) results in a smaller yet significant peak, whereas the converting activity of anterior tissues (squares) is close to the level of the control (dotted line). Radioactivity profiles shown in *b* are corrected for losses that occur during extraction and chromatography. Losses are calculated with the help of internal standard retinoic acid added before extraction and range between 40 and 55%. Elution positions of authentic retinol and retinoic acid are indicated.

METHODS. Tissue pieces (10–17) were dissected from 3–4-somite embryos (see Fig. 1c) in DMEM. Before incubating in [^3H]retinol, tissues were washed twice in L-15 medium. Tissue fragments were transferred to L-15- CO_2 medium (to 200 ml L-15- CO_2 (Speciality Media, Lavallete, NJ), add 8 ml glucose (200 mg ml $^{-1}$) and 2 ml of supplement (in Hank's balanced salt

solution) containing insulin (1 mg ml $^{-1}$), transferrin (20 mg ml $^{-1}$), sodium selenite (0.1 mg ml $^{-1}$), bovine serum albumin (1 mg ml $^{-1}$), progesterone (0.1 mg ml $^{-1}$), corticosterone (0.4 mg ml $^{-1}$), triiodothyronine, sodium salt (0.2 mg ml $^{-1}$), putrescine (3 mg ml $^{-1}$))³¹ containing 500 nM [^3H]-labelled-all-*trans*-retinol (1,369 Gbq mmol $^{-1}$, NEN; radiolabel was not further diluted with nonradioactive retinol). Tissues were incubated for 5 h at 37 °C in a CO_2 incubator (5%). They were checked for integrity at the end of incubation. Samples were chilled on dry ice, and a mixture of internal standard retinoids (17 ng all-*trans*-4-oxo-retinoic acid, 23 ng all-*trans*-5,6-epoxy-retinoic acid, 24 ng all-*trans*-3,4-didehydroretinol, 40 ng all-*trans*-3,4-didehydro retinoic acid, 22 ng 13-*cis*-retinoic acid, 41 ng all-*trans*-retinol, 49 ng retinoic acid, 27 ng all-*trans*-retinal dissolved in ethanol) and 200 μl stabilizing buffer (9 mM sodium EDTA, 25 mM sodium ascorbate, 21 mM sodium sulphate in phosphate-buffered saline, pH adjusted to 7.4) was added⁷. After brief sonication, the samples were extracted with ethylacetate/methylacetate (8:1)⁷. Extracts were first fractionated on a C_{18} HPLC column (Novapak, Waters) developed at 1 ml min $^{-1}$ with $\text{CH}_3\text{CN}:\text{CH}_3\text{OH}:\text{H}_2\text{O}$ (20:20:60) acetic acid (58:17:25). Fractions (1.2 ml) were collected and 100- μl aliquots were counted in a scintillation counter and fractions coeluting with internal standard retinoic acid were pooled, reextracted and rechromatographed on the same column. DNA was measured in the aqueous phase of the tissue extract using fluorimetry with Hoechst 33258 dye³².

FIG. 3 Model illustrating a possible mechanism of action of retinoic acid in the node. Dorsal aspect of the node. According to this model the node consists of two populations of cells, those migrating into it from the primitive streak (curved arrows) and those which are born in it^{20–23}. Mitotic activity studies^{22,23} and dye labelling in the chick²⁰, raise the possibility that the node contains a blastema-like proliferative zone. We propose that retinoic acid is continuously synthesized in the node and diffuses into the surrounding tissue (small arrows). Within the node, the product of retinoic acid concentration and time increases with time so that cells that leave the node early will have been exposed to less retinoic acid for a shorter time than cells leaving later. Note that cells exit the node in an elongated array that gives rise to midline structures (notochord, floor plate and midline endoderm) and to the medial half of somites^{20,21}. One requirement of this model is that cells 'remember' how much retinoic acid they (and their ancestors) have been exposed to. This cellular memory could involve chromosomal 'imprinting' of Hox genes in cells as they leave the node³³. The sites of retinoic acid source and proliferative zone shown are hypothetical. As discussed in the text, this model does not take into account very early primitive streak stages, before a discrete node has developed.



shifts the anterior limits of Hox genes more rostrally^{16,17,19}. In such treated embryos, cells will therefore express a particular combination of Hox genes that is normally found more posteriorly. This altered pattern may explain the homeotic transformation and posteriorization of vertebrae and rhombomeres seen after retinoic acid treatment. These observations, and our finding that the murine Hensen's node is a high point of both duplicating activity in the chick limb assay and retinoic acid synthesis, suggest that retinoic acid has a role in establishing the anteroposterior body axis.

Studies in the chick suggest that cells first ingress the node and subsequently exit as a narrow array of cells lying along the midline (for example notochord, floor plate, midline endoderm^{20,21}). In addition, there is evidence that the node is also a blastema-like tissue producing cells contributing to the midline^{20,22,23}. These observations, and our finding that the node is a high point of retinoic acid synthesis, lead us to propose the model outlined in Fig. 3. This model is consistent, for example, with a sequential 3' → 5' activation of Hox genes along the primary body axis and the fact that in cultured embryonic cells, the expression of 3' Hox genes is induced by retinoic acid more rapidly and at a lower retinoic acid concentration than genes more 5'²⁴⁻²⁶. In terms of the model, a short exposure of cells to retinoic acid in the node or early primitive streak would first activate transcription of the genes most 3' and successively longer exposures would turn on genes more 5'. This is exactly what is observed in the embryo^{27,28}. Our model is not meant to imply that retinoic acid from the node is the sole signal required for anteroposterior axial patterning, but that it could be part of a network of interacting factors, including polypeptide signalling molecules, which have also been proposed to act as morphogens during this process²⁹. If several factors are required, it will be critical to know precisely when and where they are generated, and how they interact in relation to the node. With regard to retinoic acid, it will be important to determine precisely when embryos begin to synthesize this molecule and for how long synthetic activity persists in the development of the primary axis. Preliminary studies show that conversion of retinol to retinoic acid can be detected in intact early streak stage embryos, but whether this activity is already localized to the primitive streak, remains to be determined. □

Received 24 April; accepted 7 August 1992.

- Waddington, C. H. *The Epigenetics of Birds* (Cambridge Univ. Press, Cambridge, 1952).
- Hornbruch, A. & Wolpert, L. *J. Embryol. exp. Morph.* **94**, 257-265 (1986).
- Wagner, M., Thaller, C., Jessell, T. & Eichele, G. *Nature* **345**, 819-822 (1990).
- Tickle, C. *et al. Nature* **298**, 564-566 (1982).
- Summerbell, D. *J. Embryol. exp. Morph.* **78**, 269-289 (1983).
- Eichele, G. *Trends Genet.* **5**, 246-251 (1989).
- Thaller, C. & Eichele, G. *Nature* **345**, 815-819 (1990).
- Thaller, C. & Eichele, G. *Development* **103**, 473-483 (1988).
- Ruberte, E., Dollé, P., Chambon, P. & Morriss-Kay, G. *Development* **111**, 45-60 (1991).
- Smith, S. M. & Eichele, G. *Development* **111**, 245-252 (1991).
- Rossant, J. *et al. Genes Dev.* **5**, 1333-1344 (1991).
- Durston, A. J. *et al. Nature* **340**, 140-144 (1989).
- Sive, H. L., Draper, B. W., Harland, R. M. & Weintraub, H. *Genes Dev.* **4**, 932-942 (1990).
- Ruiz i Altaba, A. & Jessell, T. *Genes Dev.* **5**, 175-187 (1991).
- Sharpe, C. R. *Neuron* **7**, 239-247 (1991).
- Kessel, M. & Gruss, P. *Cell* **67**, 89-104 (1991).
- Morriss-Kay, G. M., Murphy, P., Hill, R. E. & Davidson, D. R. *EMBO J.* **10**, 2985-2995 (1991).
- Sundin, O. H. & Eichele, G. *Development* **114**, 841-852 (1992).
- Conlon, R. A. & Rossant, J. *Development* (in the press).
- Selleck, M. A. J. & Stern, C. D. *Development* **112**, 615-626 (1991).
- Schoenwolf, G. C., Garcia-Martinez, V. & Dias, M. S. *Dev. Dynamics* **193**, 235-248 (1992).
- Lawson, K. A. & Pedersen, R. A. in *Post Implantation Development in the Mouse, Ciba Foundation Symp.* **165** (eds Chadwick D. J. & Marsh, J.) 3-26 (Wiley, Chichester, 1992).
- Stern, C. D. *Anat. Embryol.* **156**, 319-329 (1979).
- Simeone, A. *et al. Nature* **346**, 763-766 (1991).
- Simeone, A. *et al. Mech. Dev.* **33**, 215-228 (1991).
- Papalopulu, N., Lovell-Badge, R. & Krumlauf, R. *Nucleic Acids Res.* **19**, 5497-5506 (1991).
- Murphy, P. & Hill, R. E. *Development* **111**, 61-74 (1991).
- McGinnis W. & Krumlauf, R. *Cell* **68**, 283-302 (1992).
- Jessell, T. M. & Melton, D. A. *Cell* **68**, 257-270 (1992).
- Tickle, C., Summerbell, D. & Wolpert, L. *Nature* **254**, 199-202 (1975).
- Romijn, H. J. *et al. Devl Brain Res.* **2**, 583-589 (1982).
- Labarca, C. & Paigen, K. *Analyt. Biochem.* **102**, 344-352 (1980).
- Gaunt, S. J. & Singh P. B. *Trends Genet.* **6**, 208-212 (1990).

ACKNOWLEDGEMENTS. This work was funded by grants from the NIH (to B.H. and G.E.) and The McKnight Endowment Fund for Neuroscience (G.E.).

Cell-to-cell spread of calcium signals mediated by ATP receptors in mast cells

Yuri Osipchuk & Michael Cahalan

Department of Physiology and Biophysics, University of California, Irvine, California 92717, USA

RAT basophilic leukaemia cells, like mast cells from which they are derived, have surface Fcε receptors that trigger secretion of inflammatory mediators when crosslinked. Both GTP-binding proteins and a rise in cytosolic calcium concentration ($[Ca^{2+}]_i$) are implicated in the secretory mechanism¹⁻⁵. Here we use a video-imaging technique to report that transient rises in $[Ca^{2+}]_i$ initiated in an individual cell can spread from cell to cell in a wave-like pattern by means of a secreted intermediate, in the absence of gap-junctional communication. We find that the leukaemia cells, peritoneal mast cells and mucosal mast cells have cell-surface P₂-type purinergic receptors that can trigger similar $[Ca^{2+}]_i$ transients. We provide evidence that ATP is rapidly released, and that it can amplify $[Ca^{2+}]_i$ signals and initial secretory responses during antigen-stimulation of rat basophilic leukaemia cells.

As reported previously for mast cells^{3,6}, dialysis of a single rat basophilic leukaemia (RBL-2H3) cell with a patch pipette containing GTP-γS results in a rapid rise in $[Ca^{2+}]_i$ (Fig. 1a). But GTP-γS evoked an unexpected rise in $[Ca^{2+}]_i$ not only in the perfused cell but also in neighbouring cells. Although elevation of $[Ca^{2+}]_i$ in the perfused cell was sustained, it was spiky in the adjacent cells, and tended to spread radially from the patched cell. A more vigorous wave-like spread of the rise in $[Ca^{2+}]_i$ could be elicited by mechanical perturbation of a single cell in a cluster, using a patch pipette without GTP-γS (Fig. 1b). Such waves often spread from cell to cell at a velocity of 5-10 μm s⁻¹ throughout the whole microscopic field of view (270 × 320 μm; Fig. 1c). Similar spreading $[Ca^{2+}]_i$ transients could also be elicited in normal rat peritoneal mast cells cultured at high density (data not shown).

In the absence of gap-junctional coupling (Fig. 1 legend), either mechanical stimulation or intracellular GTP-γS may cause RBL or mast cells to release a substance that can induce a rise in $[Ca^{2+}]_i$ in neighbouring cells. Consistent with this hypothesis, rapidly flowing solutions inhibited the $[Ca^{2+}]_i$ signal that was initiated by mechanical stimulation (Fig. 2a). Mast cell secretory granules contain a variety of biologically active mediators, including very high concentrations of histamine^{7,8}. But bath-applied histamine had little or no effect on $[Ca^{2+}]_i$ in RBL cells at concentrations of up to 1 mM.

Another candidate substance is ATP, known to be present in mast-cell secretory granules⁹. We discovered that extracellular application of ATP to RBL cells, mucosal mast cells, and peritoneal mast cells, at concentrations from 0.5 to 10 μM, rapidly induced large transient increases in $[Ca^{2+}]_i$, lasting for 10-20 s (Fig. 2b). These responses were blocked by the P₂-receptor antagonist, suramin. ATP-induced $[Ca^{2+}]_i$ transients displayed all-or-none character, with amplitudes (averaging 0.6-1.5 μM above a resting level of about 100 nM) being virtually independent of ATP concentration. The kinetic properties, desensitization, spectrum of agonists (Fig. 2 legend), lack of Ca²⁺ or Mg²⁺ dependence (data not shown), and antagonism by suramin are consistent with a P₂ purinergic receptor-mediated signal transduction mechanism, as observed in several cell types^{10,11}. Furthermore, bath application of either a desensitizing concentration of ATP (Fig. 2c) or suramin (10 μM; data not shown) inhibited the spreading $[Ca^{2+}]_i$ rise initiated by mechanical stimulation of a single cell, suggesting that ATP receptors mediate the spreading of $[Ca^{2+}]_i$ waves.

Crosslinking of surface Fcε receptors induces heterogeneous,

slowly and was sustained. These experiments indicate that RBL cells release ATP shortly after the start of antigen-induced $[Ca^{2+}]_i$ signalling, resulting in local ATP concentration of 1–10 μM . The release of ATP by RBL cells or mast cells during stimuli that elicit secretion was corroborated by direct measurement of ATP concentration using a luciferin-luciferase assay (Fig. 4b).

Our results have directly shown that ATP is released in response to secretory stimuli in RBL and mast cells. The local extracellular ATP concentration attains levels sufficient to activate $[Ca^{2+}]_i$ responses mediated by high-affinity P_2 receptors in neighbouring cells. Elevation of $[Ca^{2+}]_i$ by Ca^{2+} ionophores can initiate secretion from mast cell lines^{14,15}. The synergistic effects of P_2 and $Fc\epsilon$ receptors on $[Ca^{2+}]_i$ signals (Fig. 3) indicate that the release of ATP can introduce an element of positive feedback in the initial secretory response. Extracellular ATP may accelerate the release of granules containing additional ATP by triggering $[Ca^{2+}]_i$ signals in quiescent neighbouring cells and in cells that have begun to degranulate. We propose that ATP receptors amplify the initial secretory response in cell clusters using this mechanism. We anticipate that these receptors would become activated at sites of tissue injury, where ATP can reach concentrations of 20 μM (ref. 16) and may thereby contribute to inflammatory responses. □

Received 22 June; accepted 4 August 1992.

1. Fernandez, J. M., Neher, E. & Gomperts, B. D. *Nature* **312**, 453–455 (1984).
2. Howell, T. W., Cookcroft, S. & Gomperts, B. D. *J. Cell Biol.* **105**, 191–197 (1987).
3. Neher, E. *J. Physiol., Lond.* **395**, 193–214 (1988).
4. Ali, H., Collado-Escobar, D. M. & Beaven, M. A. *J. Immun.* **143**, 2626–2633 (1989).
5. De Matteis, M. A., Di Tullio, G., Buccione, R. & Luini, A. *J. Biol. Chem.* **266**, 10452–10460 (1991).
6. Almers, W. & Neher, E. *FEBS Lett.* **192**, 13–18 (1985).
7. Alter, S. C. & Schwartz, L. B. *Biochim. biophys. Acta* **991**, 426–430 (1989).
8. Johnson, R. G., Carty, S. E., Fingerhood, B. J. & Scarpa, A. *FEBS Lett.* **120**, 75–79 (1980).
9. Uvnas, B. *Life Sci.* **14**, 2355–2366 (1974).
10. Burnstock, G. *Ann. N. Y. Acad. Sci.* **603**, 1–17 (1990).
11. Dubyak, G. R. & Cowen, D. S. *Ann. N. Y. Acad. Sci.* **603**, 227–243 (1990).
12. Millard, P. J., Ryan, T. A., Webb, W. W. & Fewtrell, C. *J. Biol. Chem.* **264**, 19730–19739 (1989).
13. Cheek, T. R. et al. *J. Cell Biol.* **109**, 1219–1227 (1989).
14. Cochrane, D. E. & Douglas, W. W. *Proc. natn. Acad. Sci. U.S.A.* **71**, 408–412 (1974).
15. Plaut, M. et al. *Nature* **339**, 64–67 (1989).
16. Born, G. V. R. & Kratzer, M. A. *J. Physiol., Lond.* **354**, 419–429 (1984).
17. Cornell-Bell, A. H., Finkbeiner, S. M., Cooper, M. S. & Smith, S. J. *Science* **247**, 470–473 (1990).
18. Charles, A. C., Merrill, J. E., Dirksen, E. R. & Sanderson, M. J. *Neuron* **6**, 983–992 (1991).
19. Sanderson, M. J., Charles, A. C. & Dirksen, E. R. *Cell Regul.* **1**, 585–596 (1990).
20. McCloskey, M. A. & Cahalan, M. D. *J. gen. Physiol.* **95**, 205–227 (1990).
21. Broide, D. H., Metcalfe, D. D. & Wasserman, S. I. *J. Immun.* **141**, 4298–4305 (1988).
22. Levi-Strauss, F. et al. *J. Immun.* **135**, 3454–3462 (1985).
23. Grissmer, S., Lewis, R. S. & Cahalan, M. D. *J. gen. Physiol.* **99**, 63–84 (1992).
24. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch.* **391**, 85–100 (1981).
25. Leff, P., Wood, B. E. & O'Connor, S. E. *Br. J. Pharmac.* **101**, 645–649 (1990).
26. Tatham, P. E. R., Cusack, N. J. & Gomperts, B. D. *Eur. J. Pharmac.* **147**, 13–21 (1988).
27. Tatham, P. E. R. & Lindau, M. *J. gen. Physiol.* **95**, 459–476 (1990).
28. Woldemussie, E., Maeyama, K. & Beaven, M. A. *J. Immun.* **137**, 1674–1680 (1986).
29. Chakravarty, N. *Acta pharmac. tox.* **47**, 223–235 (1980).

ACKNOWLEDGEMENTS. We thank L. Forrest for assistance in cell preparation, M. McCloskey for the RBL-2H3 cell line, reagents and advice, A. Surprenant for suggesting and providing suramin, and R. Davis for comments on the manuscript. This work was supported by an NIH grant to M.D.C.

A vertebrate actin-related protein is a component of a multisubunit complex involved in microtubule-based vesicle motility

James P. Lees-Miller*, David M. Helfman* & Trina A. Schroer†

* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

† Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218, USA

ACTIN is a cytoskeletal protein which is highly conserved across eukaryotic phyla. Actin filaments, in association with a family of myosin motor proteins, are required for cellular motile processes as diverse as vesicle transport, cell locomotion and cytokinesis^{1,2}. Many organisms have several closely related actin isoforms^{3,4}. In addition to conventional actins, yeasts contain actin-related proteins that are essential for viability^{5,6}. We show here that vertebrates also contain an actin-related protein (actin-RPV). Actin-RPV is a major component of the dynactin complex, an activator of dynein-driven vesicle movement^{7,8}, indicating that unlike conventional actins which work in conjunction with myosin motors, actin-RPV may be involved in cytoplasmic movements via a microtubule-based system.

At about the same time as the actin-related proteins act2 in *Schizosaccharomyces pombe*⁵ and Act2 in *Saccharomyces cerevisiae*⁶ were discovered, random sequencing of a human brain complementary DNA library revealed a cDNA (EST00370) encoding a 37-amino-acid peptide with 76% similarity to actin⁹. As vertebrate actins typically exhibit >93% amino-acid sequence identity⁴, we pursued the human actin-related protein as a possible homologue of act2. A full-length coding sequence was obtained by screening a human N-Tera teratocarcinoma library with oligonucleotide probes derived from clone EST00370. The translated sequence of 376 amino acids (Fig. 1) predicts a protein with M_r 42,616, which is similar to conventional actins, but its isoelectric point ($pI=6.2$) is significantly more basic ($pI=5.0$ – 5.2). Alignment of this sequence with those

of vertebrate muscle and cytoplasmic actins revealed about 54% amino-acid identity and 69% similarity, which is a significantly lower level of conservation than the 93% to 98% identity shared by the vertebrate actin isoforms, but still suggests a close evolutionary relationship and the likelihood of shared functional properties. To our knowledge, this is the first reported sequence of a vertebrate actin-related protein (actin-RPV). Actin-RPV is no more similar to vertebrate actins than to actins from invertebrates, fungi, plants and protozoans (comparison not shown) and may therefore represent a distinct subfamily rather than a vertebrate actin isoform. The 52%–55% amino-acid sequence identity that actin-RPV shares with a broad phylogenetic spectrum of conventional actins is greater than its identity with either *S. pombe* act2 (36%) or *S. cerevisiae* act2 (42%; comparison not shown), indicating that it is also unlikely to be the vertebrate homologue of either yeast actin-related protein. An evaluation of the evolutionary relationships between conventional actins and actin-related proteins will require more sequences from this latter group.

When compared to other actins, the non-conservative amino-acid substitutions in actin-RPV are scattered throughout the protein sequence and, when mapped on the actin structure deduced by X-ray crystallography¹⁰, most are found in surface loops. Conversely, the five-stranded β -sheets of actin subdomains 1 and 3 are among the most highly conserved structural features, giving an overall impression that the core structures are highly similar between conventional actin and actin-RPV. Actin-RPV probably binds ATP because 14 of the 15 residues that contact ATP or Ca^{2+} are conserved (Fig. 1). The exception is a substitution of leucine for lysine at residue 336, a position that seems to accommodate any bulky side chain as viability is not affected by engineered replacements with glutamic acid or methionine in *S. cerevisiae* actin¹¹. Several regions of actin that are believed to be involved in the intermolecular contacts necessary for actin polymerization¹² are conserved in actin-RPV, although substitutions such as lysine for Ser 323 and aspartic acid for Thr 324 may prevent copolymerization of actin-RPV with conventional actins. The sites that interact with the myosin head have diverged considerably in actin-RPV, but amino acids 361 to 364 (Glu-Tyr-Asp-Glu), believed to interact with the myosin alkaline light chain¹⁰, are conserved.

Although it remains unproven whether actin-RPV can interact with myosin or conventional actins, we have evidence for

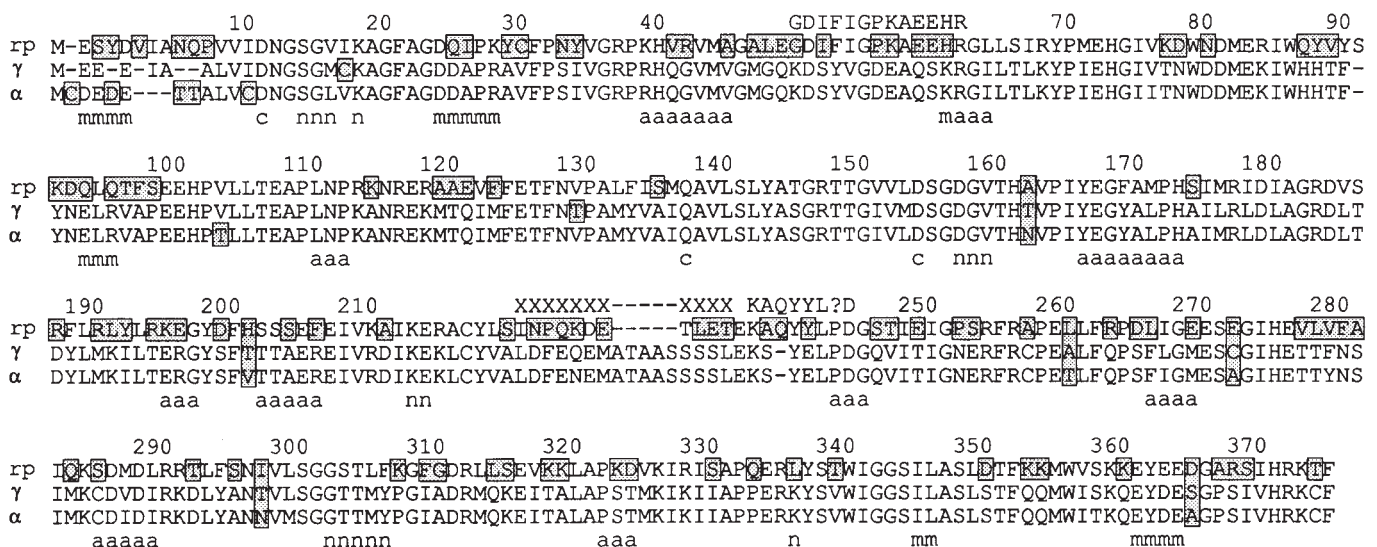


FIG. 1 Comparison of the amino-acid sequences (one-letter code) of human actin-RPV (rp), human cytoplasmic γ -actin¹⁸ and rabbit skeletal muscle α -actin¹⁹. Alignments were made with FASTDB using a McLachlan structure genetic matrix. Dashes indicate gaps required for alignment. Residues that are not similar to the consensus are shaded. Similarity is defined as (E/D), (N/Q), (K/R), (S/T), (Y/F) and (L/I/M/V). The residues in actin that interact with the nucleotide ATP (n), Ca^{2+} (c) and myosin (m) and which are involved in actin-actin interactions (a), as determined by X-ray crystallography and other methods, are indicated below the actin sequences^{10,12}. The position of the synthetic peptide CSH260 used for antibody production is indicated by XXX. The sequences of two V8 peptides from chicken brain actin-RPV are indicated above the human actin-RPV sequence. An ambiguous residue in these peptides is indicated by a question mark.

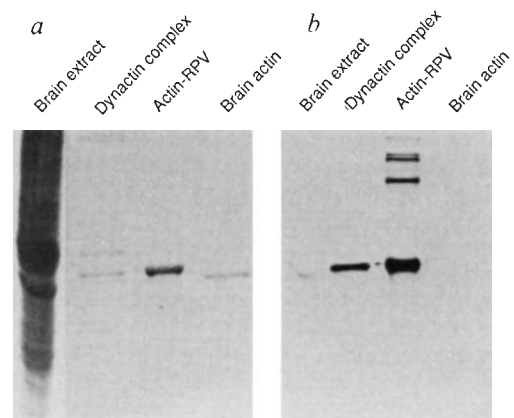
METHODS. To determine the primary structure of actin-RPV, clone EST00370 was purchased from the ATCC and sequenced with Sequenase (US Biochem.). The cDNA insert was 777 nucleotides long. We used oligodeoxynucleotides

Hact.4 (5'-AGAGCTGTACCTATCC-3') and Hact.5 (5'-ACAGTCTCTCTGA-GGTG-3'), which bracketed the 375 nucleotides at the 5' end of clone EST00370 that encoded a sequence with similarity to actin, in a polymerase chain reaction (PCR)²⁰. This was to verify that a cDNA library made from a human N-Tera teratocarcinoma contained an actin-RPV clone and to synthesize a radiolabelled probe to screen this library by plaque hybridization. PCR was carried out according to Perkin-Elmer/Cetus for library identification and modified as in ref. 21 for probe synthesis. Clone pART4 (1,679 nucleotides long and containing 48 nucleotides of 5'-untranslated sequence and 504 nucleotides of 3'-untranslated sequence) was translated to give the amino-acid sequence of actin-RPV. Chicken brain actin-RPV (purified from the dynactin complex by SDS-PAGE) was digested²² with V8 protease (Calbiochem). The resulting peptides were transferred to polyvinylidene fluoride (PVDF) membranes and sequenced on an Applied Biosystems sequencer.

its involvement in microtubule-based intracellular membrane transport. As we show later, actin-RPV is the major component of the dynein activator (dynactin) complex, which stimulates vesicle translocation driven by cytoplasmic dynein *in vitro*^{7,8}. Using an *in vitro* assay that measures the movement of membrane vesicles on microtubules, we have identified a number of cytoplasmic components that participate in vesicle translocation^{7,13,14}. Cytoplasmic dynein, an energy-dependent micro-

tubule-based motor protein that promotes movement towards microtubule minus ends, is necessary but not sufficient for minus-end-directed vesicle transport⁷. Movement is activated by the dynactin complex which consists of the 160K dynactin subunit, plus polypeptides of 62, 50, 45, 37 and 32K, the 45K component being the most abundant, present in ~8 copies per complex^{7,8}. All of these polypeptides cosediment at 20S when purified dynactin complex is sedimented into a sucrose gradient,

FIG. 2 The most abundant component of the dynactin complex is actin-RPV. **a**, Coomassie blue-stained SDS-polyacrylamide (10%) gel of (from left to right) an SDS-detergent extract of rat brain, chicken brain dynactin complex, human actin-RPV expressed in *Escherichia coli*, and chicken brain actin. **b**, Immunoblot of **a**, using a rabbit antibody against a peptide from human actin-RPV. The rat brain extract was made by homogenizing tissue in 2 volumes of 2× sample buffer²³. Dynactin complex and actin were prepared from chicken brain as described⁷. To prepare human actin-RPV, the pART4 insert containing the coding sequence for actin-RPV and the 3'-untranslated sequence was subcloned into the PET8C vector for T7 polymerase-based expression in *E. coli*²⁴. Recombinant actin-RPV was insoluble in salt solutions so that after four washes with a solution containing 25 mM Tris, 100 mM NaCl and 1 mM dithiothreitol, the pellet containing actin-RPV was extracted with this solution containing 6 M urea and 100 mM β -mercaptoethanol. The extract was mixed with an equal volume of 2× gel sample buffer, electrophoresed on an SDS-polyacrylamide gel and eluted from the gel with a BioRad transfer apparatus into 50 mM NH_4HCO_3 , 0.1% SDS. It was then freeze-dried and resuspended in sample buffer for PAGE. The oxidizing conditions of this elution step give rise to high- M_r oligomers of actin-RPV which were detected in the immunoblot. For antibody production, a peptide corresponding to amino acids 226 to 236, a region of actin-RPV divergent from actin (INPQKDETLET; see Fig. 1), preceded by a Cys residue was synthesized and coupled to keyhole limpet haemocyanin using the *m*-maleimidobenzoyl-*N*-hydroxysuccinimide ester (MBS) crosslinking reagent as directed by Pierce Co. A rabbit was injected subcutaneously at 3-week



intervals with 1 mg conjugate and was bled nine days after the fourth injection. Antibodies were affinity-purified on nitrocellulose strips containing recombinant actin-RPV as described²⁵. Antibodies were eluted with 0.2 M glycine, pH 2.8, and neutralized with 1.5 M Tris, pH 9.5. Immunoblotting was done according to ref. 24.

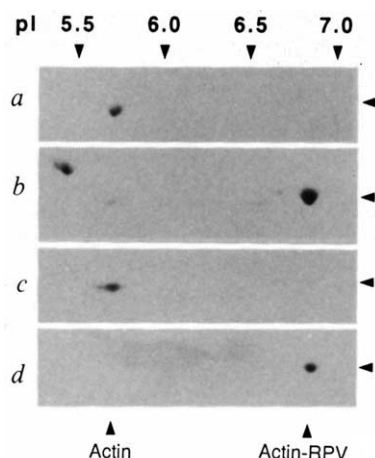


FIG. 3 Two-dimensional gel analysis of the dynactin complex. In all panels only the 40–50K region is shown. (The migration position of a 45K M_r marker is arrowed.) a, Coomassie blue-stained gel of chick embryo brain actin; b, Coomassie blue-stained gel of dynactin complex; c and d, immunoblots of dynactin complex with antibodies against actin²⁶ (c) and actin-RPV (d); different exposures are shown for the two blots. Neither antibody recognized any other proteins on the blot. Actin and dynactin complex were purified as described⁷. Two-dimensional electrophoresis was done as described²⁷ by Kendrick Labs (Madison, WI) using 10% SDS-PAGE in the second dimension. Where indicated, proteins were transferred to PVDF membrane. Immunoblotting was done as described²⁵ using the Tropix (Bedford, MA) Western Light system.

and they are coprecipitated with antibodies against the 160K dynactin subunit; these properties are indicative of a stable multiprotein assembly⁸.

A protein that resembles conventional actins in its physical and functional properties is associated with inner-arm dyneins from *Chlamydomonas axonemes*^{15,16}. Immunoblots with an anti-actin antibody (which does not crossreact with actin-RPV) indicate that a conventional actin is also present in the dynactin complex⁷. A 10% polyacrylamide gel of this complex resolves the 45K region into a minor component of higher mobility which comigrates with brain actin, and a major component of lower mobility which comigrates with bacterially expressed human actin-RPV (Fig. 2a). Moreover, this major component reacts strongly with an antibody directed against a unique actin-RPV peptide (Fig. 2b). A two-dimensional immunoblot shows the major 45K component, which is recognized by the actin-RPV specific antibody, to have a pI of 6.8 (Fig. 3). This is significantly more basic than the pI of 5.7 for brain actin, which is consistent with the isoelectric points predicted from the sequences of actin-RPV and conventional actins. Confirmation that the major 45K subunit of the chicken dynactin complex is actin-RPV was obtained by sequencing peptides generated by digestion with V8 protease. Partial sequences of two internal 16K peptides (not seen in digests of conventional actin) were identical in 20 of 21 positions with sequences in actin-RPV that diverge from conventional actin (Fig. 1).

The finding of a functional relationship between actin- and microtubule-based motility is not unprecedented. In yeast, where targeted vesicle movement is believed to use a myosin motor (MYO2), transport defects caused by a mutation in the motor can be suppressed by overexpression of a kinesin-like protein¹⁷. Our findings further support the involvement of actin and actin-related proteins in microtubule-based transport. □

5. Lees-Miller, J. P., Henry, G. & Helfman, D. M. *Proc. natn. Acad. Sci. U.S.A.* **89**, 80–83 (1992).
6. Schwob, E. & Martin, R. P. *Nature* **355**, 179–182 (1992).
7. Schroer, T. A. & Sheetz, M. P. *J. Cell Biol.* **115**, 1309–1318 (1991).
8. Gill, S. R. *et al.* *J. Cell Biol.* **115**, 1639–1650 (1991).
9. Adams, M. D. *et al.* *Science* **252**, 1651–1656 (1991).
10. Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. *Nature* **347**, 37–44 (1990).
11. Johannes, F. & Gallwitz, D. *EMBO J.* **10**, 3951–3958 (1991).
12. Holmes, K. C., Popp, D., Gebhard, W. & Kabsch, W. *Nature* **347**, 44–49 (1990).
13. Schroer, T. A., Schnapp, B. J., Reese, T. S. & Sheetz, M. P. *J. Cell Biol.* **107**, 1785–1792 (1988).
14. Schroer, T. A., Steuer, E. R. & Sheetz, M. P. *Cell* **56**, 937–946 (1989).
15. Piperno, G. & Luck, D. J. *J. Biol. Chem.* **254**, 2187–2190 (1979).
16. Piperno, G. & Luck, D. J. *Cell* **27**, 331–340 (1981).
17. Lillie, S. H. & Brown, S. S. *Nature* **356**, 358–361 (1992).
18. Erba, H. P., Gunning, P. & Kedes, L. *Nucleic Acids Res.* **14**, 5275–5294 (1986).
19. Vandekerckhove, J. & Weber, K. *Eur. J. Biochem.* **90**, 451–462 (1978).
20. Saikai, R. K. *et al.* *Science* **239**, 487–491 (1988).
21. Kelly, K. A., Stamm, S. & Kozak, C. A. *Genomics* **13**, 381–388 (1992).
22. Cleveland, D. W. *Meth. Enzym.* **96**, 222–229 (1983).
23. Laemmli, U. K. *Nature* **227**, 680–685 (1970).
24. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. *Meth. Enzym.* **185**, 60–89 (1990).
25. Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1988).
26. Lessard, J. L. *Cell Motil. Cytoskel.* **10**, 349–362 (1988).
27. O'Farrell, P. H. *J. Biol. Chem.* **250**, 4007–4021 (1975).

ACKNOWLEDGEMENTS. We thank S. Stamm for help with PCR, J. Skowronski for the teratocarcinoma library, and M. Eckley, D. Hammond and A. Salcedo for technical assistance. This work was supported by the NIH (D.M.H. and T.A.S.), the ACS (J.P.L.), and the Packard Foundation and Searle Scholar's Program (T.A.S.). D.M.H. is an Established Investigator of the American Heart Association.

Centractin is an actin homologue associated with the centrosome

Sean W. Clark & David I. Meyer*

Department of Biological Chemistry, UCLA School of Medicine and The Molecular Biology Institute, Los Angeles, California 90024, USA

ACTIN is one of the most ubiquitous, abundant and well-conserved proteins of eukaryotes, participating in many crucial cellular processes including the maintenance of cell shape, motility and cell division^{1,2}. Actins from the most divergent sources still share amino-acid identities in excess of 70% (ref. 3). This may well explain why low-abundance homologues of actin have been difficult to isolate. Genes encoding distant relatives of actin in budding and fission yeast have now been cloned^{3,4}. We report here the discovery of a vertebrate actin-like protein, which we name **centractin**. A full-length complementary DNA clone was isolated whose sequence reveals amino-acid identities with actin of over 50%, increasing to more than 70% when conservative amino-acid changes are considered. Northern analysis and western blotting indicate a ubiquitous tissue and species distribution. Morphological and biochemical criteria show that **centractin** is associated with centrosomes.

A monoclonal antibody (raised against a rough endoplasmic reticulum (ER)-specific membrane protein⁵), when used for expression cloning from an MDCK cell cDNA library⁶, consistently yielded fusion proteins having open reading frames with similarity to actin (for which the simplest explanation is a shared epitope between the ER-specific protein and **centractin**). The longest clone comprised 2,830 base pairs (out of a possible 2,909 base pairs (bp)), and thus spanned 97% of the messenger RNA. The RACE procedure⁷ was used to obtain the missing 3%. The principal open reading frame encodes a protein of 376 amino acids with M_r of 42,600 (42.6K; Fig. 1a). One of several actin sequences showing an overall amino-acid identity with **centractin** of >50% (from Genbank and SwissProt databases) was human α -actin at 53%. If conservative amino-acid changes are considered, **centractin** is similar to actin over more than 70% of its sequence (Fig. 1b).

Centractin's similarity to actin is not random throughout the protein. Rather, superimposition of the **centractin** sequence on the three-dimensional structure of actin indicates that regions corresponding to the core of actin are more conserved^{8,9}. This

Received 5 May; accepted 14 July 1992.

1. Cheney, R. E. & Mooseker, M. S. *Curr. Opin. Cell Biol.* **4**, 27–35 (1992).
2. Satterwhite, L. L. & Pollard, T. D. *Curr. Opin. Cell Biol.* **4**, 43–52 (1992).
3. Meagher, R. B. & McLean, B. G. *Cell Motil. Cytoskel.* **16**, 164–166 (1990).
4. Rubenstein, P. A. *Bioessays* **12**, 309–315 (1990).

* To whom correspondence should be addressed.

canine sequence near the amino terminus and 86 amino acids near the carboxy terminus. The N-terminal overlap is at least 90% identical, whereas that at the C terminus is between 95 and 100% (both values depend on unknown residues in the human clones). Comparison with a complete human centractin sequence¹¹ indicates that canine and human centractins are identical. Thus we conclude that the N-terminal human cDNA fragment used for comparison in Fig. 1b represents yet another closely related gene. Centractin, like actin, may thus have several isoforms. It is notable that complete identity between canine and human centractin exceeds even the interspecies conservation of actin. Conservation across a wider range of species is supported by immunological data (see later).

The centractin gene is expressed ubiquitously and at low levels. Northern blot analysis was used to determine the tissue distribution of centractin mRNA. Using our canine centractin cDNA, we probed poly(A)⁺ mRNA from several human tissues (Fig. 2a, upper panel). We detected a single mRNA species in each tissue of ~3 kilobases (kb), larger than the 2.2-kb β -actin mRNA. This difference in size between actin and centractin mRNAs is due almost entirely to the large 3' untranslated sequence (Fig. 1a). Blotting with α - and β -actin probes (Fig. 2a, middle and lower panels) allowed a comparison of centractin levels between tissues. In relation to the β -actin level in each tissue, the quantity of centractin was fairly consistent, although slightly lower in placental and kidney tissue. As β -actin is downregulated in muscle tissues expressing α -actin¹², simple comparison of centractin and β -actin levels cannot be made in muscle tissue. Nonetheless, the absolute level of centractin expression in skeletal muscle is high relative to other tissues. As the centractin-probed blot (Fig. 2a, upper panel) was exposed for ten times longer than the actin-probed blots (Fig. 2a, middle

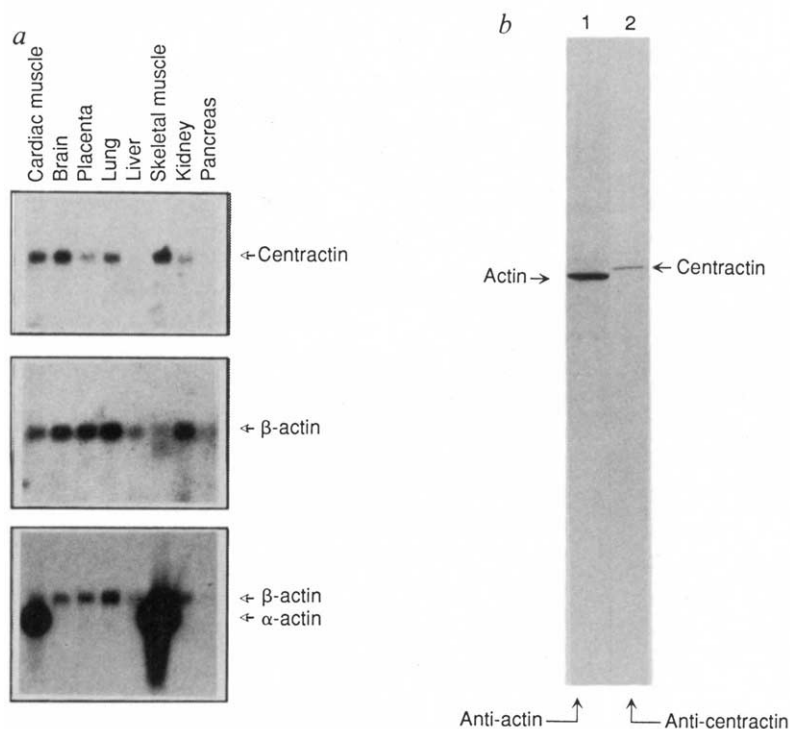
and lower panels), centractin mRNA levels must be far lower than actin mRNA in all tissues.

Centractin was biochemically and morphologically characterized using affinity-purified antibodies prepared against recombinant centractin produced in *Escherichia coli*. An immunoblot of bovine brain homogenate probed with affinity-purified rabbit anti-centractin serum is shown in Fig. 2b, lane 2. An identical tissue sample stained with anti-actin serum is shown in lane 1. Apart from a lack of cross-reactivity between anti-actin and anti-centractin antibodies, anti-centractin serum is monospecific for a protein that migrates slightly more slowly than actin on SDS-polyacrylamide gels, in agreement with the size deduced from the predicted amino-acid sequence. We consistently found centractin to be less abundant than actin in all tissues and species tested using several mouse and rabbit sera. Similar results were obtained (data not shown) on immunoblots of homogenates of rat and mouse brain, and a variety of cultured cell lysates, including canine fibroblasts and thymocytes; mouse thymocytes, hypothalamic neurons and pituitary cells; rat pheochromocytoma cells; HeLa cells and *Xenopus* oocytes.

Indirect immunofluorescence was used to determine the cellular localization of centractin. Shown for comparison is typical staining of stress fibres using anti-actin antibodies (Fig. 3a). Centractin fluorescence (Fig. 3c, d) revealed a faint cytoplasmic staining greatly concentrated at a single bright spot in the perinuclear region of each cell (Fig. 3c), although several cells had two bright perinuclear spots (Fig. 3d). Such a pattern and distribution is typical of centrosomes¹³. Immunofluorescence using affinity-purified anti-centractin serum, preincubated with an excess of recombinant centractin, resulted in the loss of centrosomal staining and a decrease in the cytoplasmic staining (Fig. 3e). As the centrosome is at the focus of an interphase

FIG. 2 Northern analysis of human centractin expression in various tissues and identification of the centractin protein. a, Single blot was probed independently with three different probes recognizing: upper panel, centractin; middle panel, β -actin; and lower panel, α - and β -actin. Approximate mRNA sizes are: centractin, 3.1 kb; β -actin, 2.2 kb; and α -actin, 1.8 kb. b, Bovine brain homogenate was immunoblotted using either rabbit anti-actin (lane 1) or affinity-purified rabbit anti-centractin (lane 2). A comparison indicates that each antibody is highly specific for its respective antigen. Note the significantly lower abundance of centractin and slightly slower migration, as predicted by its sequence.

METHODS. a, Human multiple tissue northern blot, containing 2 μ g poly(A)⁺ RNA from each tissue, was obtained from Clontech. Pre-hybridization, hybridization, washing and stripping were carried out according to the supplier's recommendations, except 6 $\times$ SSPE and no formamide were used. Hybridization conditions were: centractin probe, 50 °C for 18 h; β -actin probe, 42 °C for 16 h; α -actin probe, 42 °C for 7 h. Final washes were: centractin probe, 0.1 $\times$ SSC at 65 °C for 1 h; β -actin probe, 0.1 $\times$ SSC at 50 °C for 1 h; α -actin probe, 0.2 $\times$ SSC at 65 °C for 10 min. Blots were exposed for 34 h for centractin and 3 h for actin. Probes were labelled using a Random-Primed DNA labelling kit (USB). Gel-purified probes were: upper panel, canine centractin cDNA 740-bp *Bam*HI fragment corresponding to the C-terminal 193 amino acids plus 160 bp of 3' untranslated region; middle panel, human β -actin²⁸ 600 bp *Eco*RI/*Bam*HI fragment corresponding to the 3' UT; and lower panel, mouse α -actin²⁹ *Pst*I fragment, recognizing both α - and non-muscle actins. b, For production of antisera, an expression construct, centractin-(His)₆, was produced by adding six histidines to the C terminus of the centractin sequence using PCR. This modified centractin cDNA consisting of the entire coding region plus the C-terminal histidines was assembled into the pET8c vector and induced for high-level expression according to ref. 30. The addition of the six C-terminal histidines enabled recombinant centractin-(His)₆ to be purified to homogeneity using Ni²⁺ chromatography³¹. Purified centractin-(His)₆ was used to immunize both rabbits and mice. Sera were affinity-purified over columns consisting of centractin-(His)₆ coupled to Affigel-10 (BioRad) according to the manufacturer's suggestions. These antibodies



showed no cross-reactivity with actin (nor with other proteins containing six C-terminal histidine residues; data not shown) on either immunoblots or in immunofluorescence. Immunoblotting: equal quantities of bovine brain homogenate were applied to each lane of a 10–15% gradient gel³². Proteins were transferred to nitrocellulose and immunoblotted³³ with centractin and actin antibodies. Primary antibodies were detected using secondary antibodies conjugated to alkaline phosphatase (Cappel) using NBT/BCIP (BioRad) as substrate. Rabbit anti-actin antiserum, specific for α - and γ -smooth muscle and γ -non-muscle isoforms, was a gift from J. C. Bulinski.

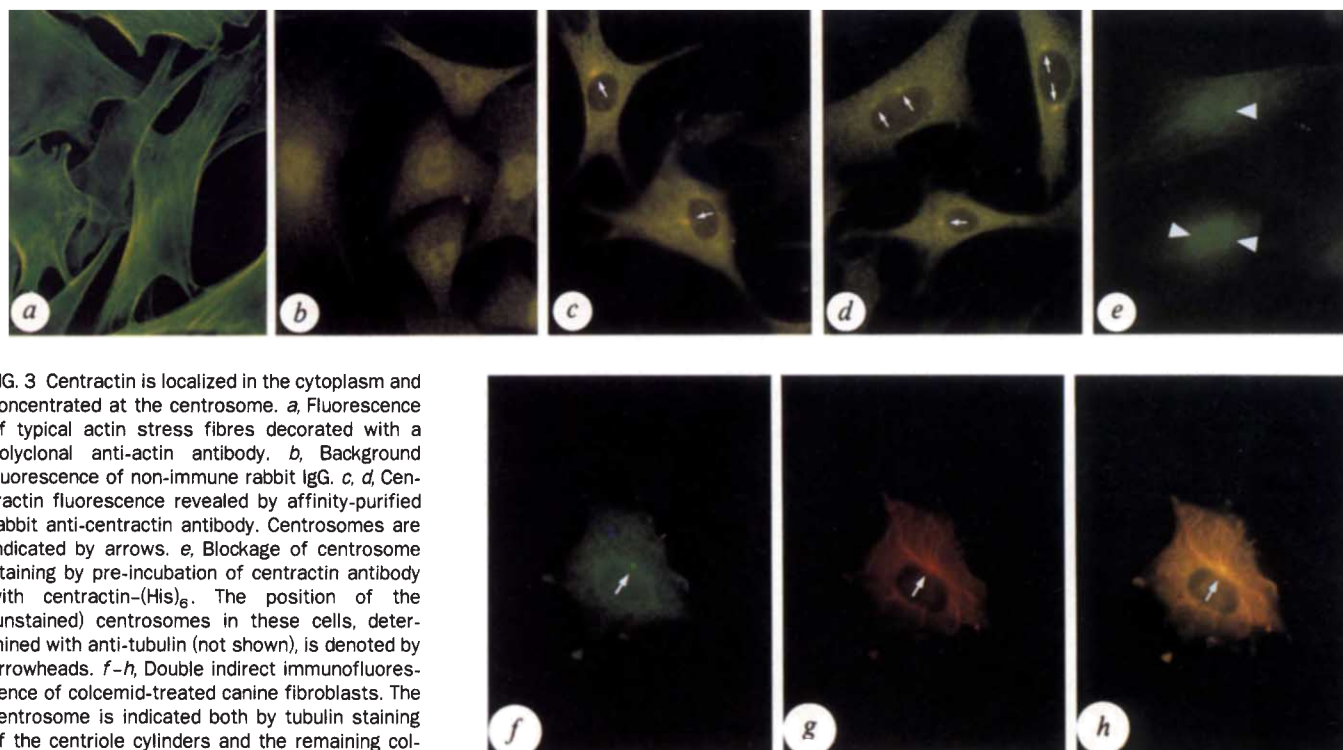


FIG. 3 Centractin is localized in the cytoplasm and concentrated at the centrosome. *a*, Fluorescence of typical actin stress fibres decorated with a polyclonal anti-actin antibody. *b*, Background fluorescence of non-immune rabbit IgG. *c*, *d*, Centractin fluorescence revealed by affinity-purified rabbit anti-centractin antibody. Centrosomes are indicated by arrows. *e*, Blockage of centrosome staining by pre-incubation of centractin antibody with centractin-(His)₆. The position of the (unstained) centrosomes in these cells, determined with anti-tubulin (not shown), is denoted by arrowheads. *f-h*, Double indirect immunofluorescence of colcemid-treated canine fibroblasts. The centrosome is indicated both by tubulin staining of the centriole cylinders and the remaining colcemid-resistant microtubules radiating from the centrosome (arrow). *f*, FITC fluorescence indicating centractin. *g*, Texas Red fluorescence indicating β -tubulin. *h*, Double exposure indicating fluorescence overlap.

METHODS. *a-h*, Canine fibroblasts were grown in DMEM medium +10% fetal bovine serum in a 5% CO₂ atmosphere on 12-mm glass coverslips. Cells were fixed in 3% formaldehyde in PBS supplemented with 10 μ M CaCl₂ and MgCl₂, quenched with 150 mM NH₄Cl in PBS and solubilized in 0.2% Triton X-100 in PBS. Fixed cells were blocked with 0.2% gelatin in PBS. Antibodies were used at 1:100 dilution for actin and 12 μ g ml⁻¹ for centractin and non-immune rabbit IgG (Cappel). Fluorescein isothiocyanate (FITC)-conjugated secondary antibodies were from Sigma. Cells were post-fixed in 3% formaldehyde and mounted in 80% glycerol/10 mM Tris-HCl, pH 8.6, containing 2% 1,4-diazabicyclo(2,2,2)octane (Kodak) to reduce fluorescence bleaching. *e*, Affinity-purified rabbit anti-centractin antibody was pre-incubated for 30 min at 4 °C with a 200-molar excess of centractin-(His)₆ before immunofluorescence. *f-h*, Canine fibroblasts were treated for 60 min with 0.5 μ g ml⁻¹ colcemid (*N*-deacetyl-*N*-methylcolchicine; Sigma) before fixation. Anti- β -tubulin monoclonal antibody (from E. Schweitzer) was used at 1:1,000. Texas Red-conjugated secondary antibody (Cappel) was used to visualize the tubulin antibody.

microtubule organizing centre (MTOC), we double-stained colcemid-treated fibroblasts with rabbit anti-centractin and mouse anti-tubulin antibodies. Treatment of cells *in situ* with colcemid results in a depolymerization of most cytoplasmic microtubules but leaves the centriole cylinders intact¹⁴. The micrographs in Fig. 3*f, g, h* show that the bright spots seen with anti-centractin antibody are indeed at the focal point of the MTOC. This indicates that localization of centractin to the centrosome does not require the presence of most of the cytoplasmic microtubules. Conversely, cytoplasmic staining is reduced in colcemid-treated cells. Decreased cytoplasmic staining in the absence of microtubules may be explained if a portion of centractin is also associated with the dynactin complex¹¹ which is involved in microtubule-based vesicle transport¹⁵.

We used cell fractionation to corroborate these morphological observations. Highly purified fractions of centrosomes were obtained by ultracentrifugation on sucrose gradients¹⁶. The fractions were characterized both morphologically and biochemically for a known centrosomal marker (human autoimmune antigen), a centriolar marker (mouse β -tubulin) and centractin. A sample of each gradient fraction, sedimented onto coverslips and assayed for centrosomes by double immunofluorescence (anti-centrosomal and anti-centractin antibodies), revealed an overlapping, bipunctate staining in fractions known to contain centrosomes^{16,17} (result not shown). Immunoblots of isolated centrosomes (Fig. 4*a*) showed that centractin distribution was coincident with centrosomes detected by centrosomal and centriolar markers. A lesser but reproducible amount of what may be unassociated centractin was seen in the upper portion of the sucrose gradients (not shown).

Isolated centrosomes were also characterized by two-dimensional electrophoresis to allow comparison with other protein profiles. There is a 'centrosome-specific' polypeptide of similar molecular mass to centractin which has an isoelectric point about one pH unit higher than actin¹⁶. We analysed our purified centrosomes by two-dimensional electrophoresis, identifying centractin by immunoblotting. Centractin migrates to the basic side of actin in the first dimension and at a slightly lower mobility than actin in the second dimension (Fig. 4*b*). This corresponds perfectly to the centrosome-specific polypeptide seen by Bornens *et al.*¹⁶ on two-dimensional gels, providing independent confirmation of centractin's centrosomal location.

A question that arises from these results concerns centractin's role in centrosomal function. A new member of the tubulin family, γ -tubulin¹⁸, has been localized to the pericentriolar material of centrosomes¹⁹ and to the spindle-pole body in yeast²⁰. Like centractin, γ -tubulin is expressed at low levels compared with abundant homologues²¹ and has been implicated in microtubule nucleation²². In this role, γ -tubulin may be a link between the microtubule array and the centrosome, a centrosomal protein through which the cytoskeleton and the centrosome interact. Likewise, centractin may link the centrosome and actin or other proteins. Interestingly, there are two linkages within the centrosome, the intercentriolar linkage¹⁶ and the centrosome-nuclear linkage²³, which could involve actin or an actin-like protein.

Consensus motifs that often specify post-translational modifications revealed the existence of centractin sequences, not present in actin, which might enable centractin to be phosphorylated by casein kinase II (ref. 24). Such modifications in

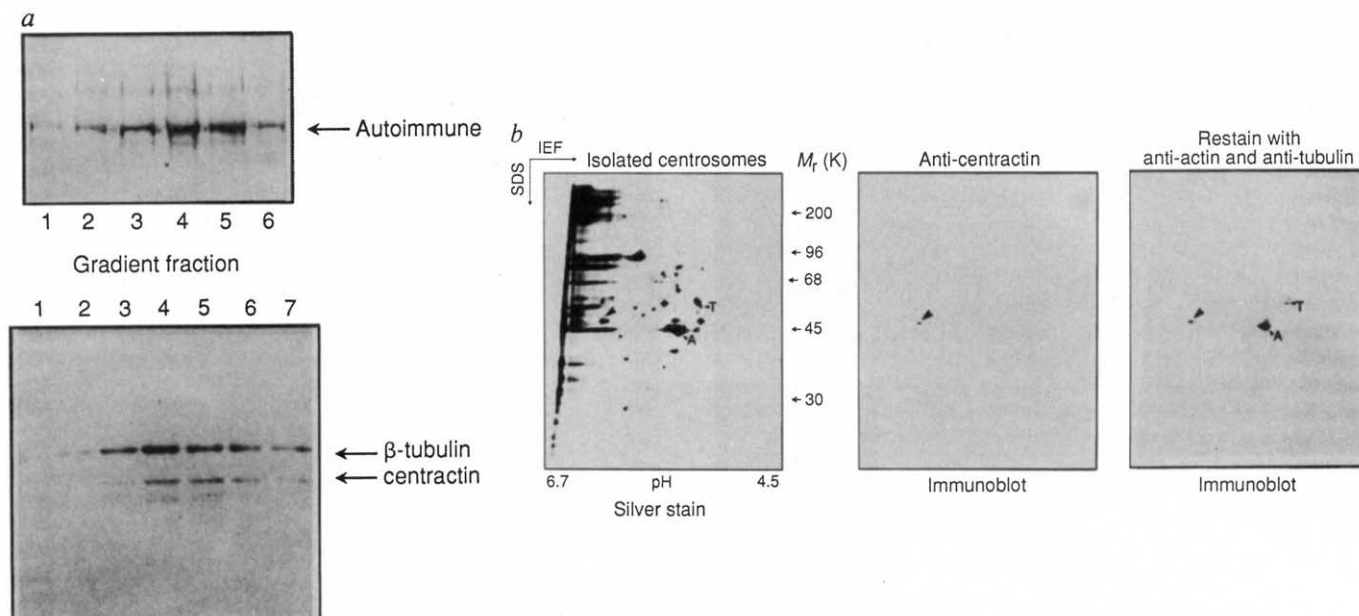


FIG. 4 Identification and two-dimensional electrophoretic analysis of centractin in purified HeLa cell centrosomes. **a**, Upper panel, immunoblot of sucrose gradients used in centrosome isolation decorated with a human autoimmune centrosomal marker which recognizes a 200K protein (gift from E. Tan). Lanes 1–6 represent fractions 1–6, collected from the bottom of the gradient. Lower panel, same fractions as in upper panel, reacted with anti-centractin and anti-tubulin antibodies. Lanes 1–6 represent fractions 1–6, lane 7 represents fraction 8. Lanes in upper panel contain 2.5 times more centrosomes as loaded in lower panel. **b**, Isoelectric focusing (IEF)/SDS-polyacrylamide gel electrophoresis analysis of isolated HeLa cell centrosomes. Left panel, silver stain of purified centrosomal proteins. Middle panel, immunoblot with anti-centractin antibody. Right panel, reprobing of the blot in middle panel with anti-actin and anti-tubulin antibodies. The large arrowhead indicates centractin. The small arrowheads indicate T for tubulin, and A for actin.

METHODS. **a**, Centrosomes were isolated using a low-ionic-strength/detergent lysis method¹⁶ from HeLa cells in suspension culture, grown in DMEM

plus 10% FBS/5% CO₂. Fractions (500 μ l) were collected from the bottom of the gradient. Upper panel, each fraction (100 μ l) was diluted with 300 μ l 10 mM K-PIPES, pH 7.2, concentrated by centrifugation onto a 20- μ l cushion of 70% sucrose in 10 mM K-PIPES, pH 7.2, using a TLA100.1 rotor at 55,000 r.p.m. for 20 min and analysed on a 10% SDS-PAGE gel. Lower panel, Each fraction (40 μ l) was analysed on a 10–15% gradient gel. Proteins were transferred to nitrocellulose and immunoblotted as for Fig. 2b. The blot shown in the lower panel was first probed with the affinity-purified rabbit anti-centractin antibody, then reprobed for β -tubulin. **b**, Purified HeLa centrosomes (250 μ l, fraction 4) were diluted and centrifuged as in **a**. First-dimension IEF gels (10 cm $\times$ 2 mm) were prepared as described¹⁶ then run and equilibrated according to ref. 34. Two first-dimension IEF gels, run in parallel, were applied to a single 10–15% gradient gel for the second dimension. One half of this gel was processed for silver staining³⁵ while the other was transferred to nitrocellulose and immunoblotted. The monoclonal actin antibody, specific for non-muscle actins, was a gift from B. Jockusch.

centractin could easily be identified and would provide a way to control centractin function or its appearance at different phases of the cell cycle.

A report of actin-like genes in yeast^{3,4} provides an interesting parallel to our discovery of centractin. Like centractin, these genes have retained a 40–50% identity with those of actin. And like centractin, the conserved residues are concentrated in domains corresponding to the core of actin. Yet, on a purely numerical basis, centractin, *Schizosaccharomyces pombe* ACT2 and *Saccharomyces cerevisiae* ACT2 are no more similar to one another than they are to actin. Whether these three proteins are functionally homologous can best be determined by genetic complementation, but considering centractin's centrosomal location, it is interesting that disruption of the ACT2 gene results in a lethal phenotype, which is indicative of a lesion late in the cell cycle³. □

- Adams, M. D. *et al.* *Science* **252**, 1651–1656 (1991).
- Lees-Miller, J. P., Helfman, D. M. & Schroer, T. A. *Nature* **359**, 244–246 (1992).
- Buckingham, M. E. *Essays Biochem.* **20**, 77–109 (1985).
- Gosti-Testu, F. *et al.* *EMBO J.* **5**, 2545–2550 (1986).
- Osborn, M. & Weber, K. *Proc. natn. Acad. Sci. U.S.A.* **73**, 867–871 (1976).
- Gill, S. R. *et al.* *J. cell. Biol.* **115**, 1639–1650 (1991).
- Bornens, M. *et al.* *Cell Motil. Cytoskel.* **8**, 238–249 (1987).
- Mitchison, T. & Kirschner, M. *Nature* **312**, 232–237 (1984).
- Oakley, C. E. & Oakley, B. R. *Nature* **338**, 662–664 (1989).
- Zheng, Y., Jung, M. K. & Oakley, B. R. *Cell* **65**, 817–823 (1991).
- Oakley, B. R., Oakley, C. E., Yoon, Y. & Jung, M. K. *Cell* **61**, 1289–1301 (1990).
- Stearns, T., Evans, L. & Kirschner, M. *Cell* **65**, 825–836 (1991).
- Joshi, H. C., Palacios, M. J., McNamara, L. & Cleveland, D. W. *Nature* **356**, 80–83 (1992).
- Maro, B. & Bornens, M. *Biol. Cell* **39**, 287–290 (1980).
- Kuenzel, E. A., Mulligan, J. A., Sommercorn, J. & Krebs, E. G. *J. biol. Chem.* **262**, 9136–9140 (1987).
- Stanley, K. K. & Luzio, J. P. *EMBO J.* **3**, 1429–1434 (1984).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: a Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, New York, 1989).
- Marchuk, D., Drumm, M., Saulino, A. & Collins, F. S. *Nucleic Acids Res.* **19**, 1154 (1991).
- Ponte, P., Gunning, P., Blau, H. & Kedes, L. *Molec. cell. Biol.* **3**, 1783–1791 (1983).
- Minty, A. J. *et al.* *J. biol. Chem.* **256**, 1008–1014 (1981).
- Studier, F. W. & Moffatt, B. A. *J. molec. Biol.* **189**, 113–130 (1986).
- Bush, G. L., Tassin, A. M., Friden, H. & Meyer, D. I. *J. biol. Chem.* **266**, 13811–13814 (1991).
- Laemmli, U. K. *Nature* **227**, 680–685 (1970).
- Towbin, H., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350–4354 (1979).
- O'Farrell, P. H. *J. biol. Chem.* **250**, 4007–4021 (1975).
- Ansorge, W. *J. biochem. biophys. Meth.* **11**, 13–20 (1985).

Received 23 June; accepted 13 August 1992.

- Pollard, T. D. A. *Rev. Biochem.* **55**, 987–1035 (1986).
- Pollard, T. D. *Curr. Opin. Cell Biol.* **2**, 33–40 (1990).
- Schwob, E. & Martin, R. P. *Nature* **355**, 179–182 (1992).
- Lees-Miller, J. P., Henry, G. & Helfman, D. M. *Proc. natn. Acad. Sci. U.S.A.* **89**, 80–83 (1992).
- Hortsch, M., Avossa, D. & Meyer, D. I. *J. cell. Biol.* **103**, 241–253 (1986).
- Herz, J. *et al.* *FEBS Lett.* **276**, 103–107 (1990).
- Frohman, M. A., Dush, M. K. & Martin, G. R. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8998–9002 (1988).
- Kabsch, W. *et al.* *Nature* **347**, 37–44 (1990).
- Holmes, K. C., Popp, D., Gebhard, W. & Kabsch, W. *Nature* **347**, 44–49 (1990).

ACKNOWLEDGEMENTS. We thank B. Jockusch, C. Bulinsky, E. Tan and E. Schweitzer for antibodies; M. Carey, E. Schweitzer, E. Neufeld, D. Wiest, B. Howard, S. Smale and P. Mellon for cell cultures; B. Winsor, D. Helfman, R. Waterston and M. Adams for access to sequence data before publication; B. Pilz for cell culture and antibody production; E. Schweitzer for technical advice and facilities; B. Dobberstein for the pTEx library; E. Ujita-Lee and M. Nussmeier for actin probes; A. Savitz and Y. Sun for advice and G. Payne for critically reading the manuscript. S.W.C. is supported by a USPHS training grant; D.I.M. acknowledges a faculty research award from the American Cancer Society. The research was supported by a grant (to D.I.M.) from the USPHS.

Escherichia coli cell-division gene *ftsZ* encodes a novel GTP-binding protein

Debabrata RayChaudhuri & James T. Park

Department of Molecular Biology and Microbiology,
Tufts University School of Medicine, Boston, Massachusetts 02111, USA

ESCHERICHIA COLI divides by forming a septum across the middle of the cell. The biochemical mechanism underlying this process is unknown. Genetic evidence suggests that of all the *fts* (filamentation temperature sensitive) genes^{1,2} involved in *E. coli* cell division, *ftsZ* plays a central role at the earliest known step of septation³⁻⁵. Here we show that FtsZ protein binds GTP *in vitro* using unusual sequence elements⁶⁻⁸. In contrast, such binding to the product of the conditional-lethal *ftsZ84* allele is impaired. Purified FtsZ displays a Mg²⁺-dependent GTPase activity which is markedly reduced in the FtsZ84 protein. FtsZ copurifies with near stoichiometric amounts of noncovalently-bound GDP, implying the presence of a GTPase cycle *in vivo*, similar to that known for signal-transducing GTP-binding proteins^{8,9}. We also show that a small fraction of FtsZ exists as a distinct membrane-associated species that binds GTP. The membrane association of FtsZ and the known ability of GTPases to act as molecular switches^{8,9} implicate FtsZ in a GTP-activated signal transduction pathway that may regulate the start of septation in *E. coli*.

Genetic studies have identified cell division *fts* genes by the isolation of temperature-sensitive mutants that are blocked in septation, but proficient in DNA replication and chromosome

segregation^{1,2}. Genetic evidence indicates a pivotal role for *ftsZ* very early in septation³⁻⁵, yet the biochemical function of the gene product has remained elusive. A search of the databases did not reveal any protein with significant homology to FtsZ, but an inspection of the FtsZ amino-acid sequence^{6,7} revealed motifs that seemed to be a variation of the consensus GTP/GDP-binding domain of the GTPase superfamily of proteins⁸ (Table 1).

To test whether FtsZ is indeed a GTP-binding protein, we overexpressed it in *E. coli* from a T7 promoter (Fig. 1a) and tested GTP-binding using the blot overlay assay¹⁰⁻¹². Many known monomeric GTP-binding proteins (*M_r* 20-30K), after SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and western transfer to a membrane, refold and bind GTP *in situ*. Because FtsZ has an apparent *M_r* of ~43K (Fig. 1a, lane 3), we included a renaturation step¹² before probing blots with GTP. After cell breakage, FtsZ was localized by immunoprecipitation to both the cytoplasmic and membrane fractions of cells expressing the protein either from the chromosomal gene or from the gene under T7-promoter control. Surprisingly, the membrane-bound form of FtsZ showed retarded mobility on SDS gels. As shown in Fig. 1b, FtsZ present in the cytoplasmic (lane 4) or the membrane fraction (lane 5) of overproducing cells could bind [α -³²P]GTP, whereas FtsZ expressed from the chromosomal copy of the gene did not reveal any such binding (lanes 2 and 3) presumably because there were insufficient amounts of FtsZ in these lanes. We did not detect [α -³²P]ATP-binding to FtsZ on blots (data not shown).

To test whether GTP-binding to FtsZ might be relevant for its cell-division function, we cloned the thermosensitive *ftsZ84* allele and overexpressed the mutant protein (Fig. 1a, lanes 5

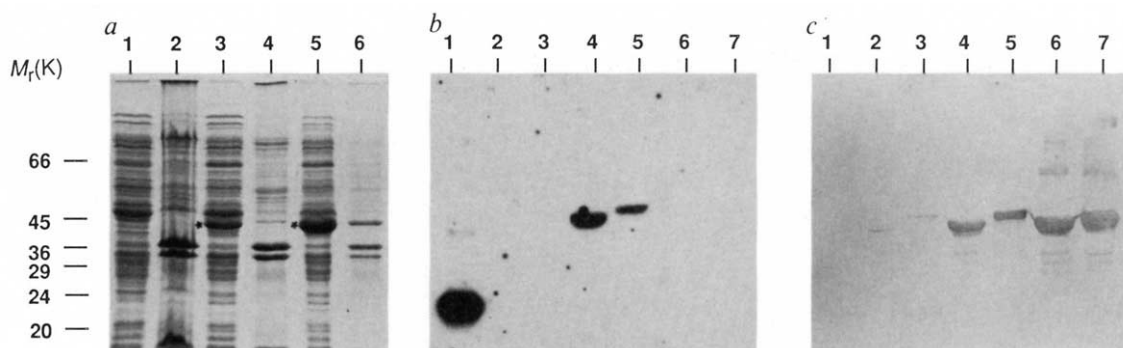


FIG. 1. Overexpression of FtsZ and FtsZ84 proteins in *E. coli* and [α -³²P]GTP-binding on western blots. **a**, Coomassie blue-stained 12% SDS-polyacrylamide gel profile of cytosolic and membrane fractions from overproducing cells; lanes 1 and 2, cytosolic and membrane fractions, respectively, from isopropyl- β -D-thiogalactoside (IPTG)-induced pET-3 control cells; lanes 3 and 4, as lanes 1 and 2, but from induced pET-3Z⁺ cells; lanes 5 and 6, as lanes 1 and 2, but from induced pET-3Z84 cells. Each lane received ~15 μ g protein. The asterisks next to lanes 3 and 5 indicate the positions of soluble, overproduced FtsZ and FtsZ84 proteins, respectively. **b**, [α -³²P]GTP-binding to a western blot; lane 1, recombinant p21 Ha-Ras protein (2 μ g); lanes 2-7, as in **a**, lanes 1-6; the amounts of protein were: lane 2, 20 μ g; lane 3, 28 μ g; lane 4, 17 μ g; lane 5, 22 μ g; lane 6, 12 μ g; lane 7, 20 μ g. **c**, Immunostaining of the western blot in **b** with anti-FtsZ polyclonal antibody.

METHODS. FtsZ and FtsZ84 proteins were overexpressed using Studier's T7 expression system³⁴. To clone *ftsZ* gene under the control of T7 gene10 promoter, pZAQ⁷ DNA was cut at the unique *Bsu*36I site located just upstream of the putative Shine-Dalgarno sequence of *ftsZ*⁶ and a *Bsu*36I-*Cla*I fragment (~1.4 kb) containing the entire *ftsZ* gene was purified. After filling out the ends with the Klenow fragment of DNA Polymerase I, the *ftsZ* fragment was inserted into the filled-out *Bam*HI ends of pET-3 (ref. 34). The resulting plasmid pET-3Z⁺ overexpressed FtsZ in *E. coli* strain BL21(DE3) (ref. 34) after induction with IPTG. The *ftsZ84* mutant allele was cloned from the genomic DNA of strain JFL100 (ref. 7) using the polymerase chain reaction (PCR)³⁵. The two PCR primers were 5'-GAAAAGCTTCGCCAACAAAGG starting three bases upstream of the unique *Hind*III site in *ftsA*, and 5'-GAGGGATCCACCCAAATCCAGTCAATTC which introduced a synthetic

*Bam*HI site (underlined) 20 bases downstream of the *ftsZ* stop codon⁶. Amplification was done for 30 cycles using *Taq* Polymerase³⁵ (1 min at 94 °C, 2 min at 55 °C, and 2 min at 72 °C; final extension for 10 min at 72 °C). The PCR product (~1.5 kb) was digested with *Bsu*36I and *Bam*HI, and inserted into the *Bam*HI site of pET-3 by blunt-end ligation to generate pET-3Z84. The cloned *ftsZ84* gene was sequenced using Sequenase 2.0 (USB) to ensure that the only mutation was the Gly 105 to Ser change⁷. Sequencing the wild-type *ftsZ* gene in parallel revealed one silent change (codon 327, ACG to ACC), in addition to the corrections reported⁷. pET-3Z84 directed the synthesis of high levels of FtsZ84 in BL21(DE3) at 30 or 37 °C on IPTG induction. Proteins were overproduced in cells grown in M9-ZB-0.2% glucose-100 μ g ml⁻¹ ampicillin medium following 2.5 h IPTG (0.5 mM) induction³⁴. Cell lysis and subcellular fractionation were as ref. 36, except that spheroplasts were resuspended in 10 mM Tris-HCl, pH 8.0, 5 mM MgCl₂, 200 μ g ml⁻¹ phenylmethylsulphonyl fluoride (PMSF), 2 μ g ml⁻¹ leupeptin before sonication, and the cytosolic and total membrane fractions separated by ultracentrifugation (>300,000g). The membrane fraction was washed and resuspended in the above buffer. Western transfer of SDS-polyacrylamide gels and blot assay with [α -³²P]-labelled GTP or ATP were done essentially as in ref. 12. The GTP-binding buffer was 50 mM Tris-HCl, pH 7.5, 100 mM potassium acetate, 5 mM MgCl₂, 1 mM EDTA, 1 mM dithiothreitol (DTT) and 0.3% (v/v) Tween-20. For immunostaining of western blots, polyclonal rabbit antibodies raised against FtsZ (manuscript in preparation) were used at 1:15,000 dilution and the blot developed with goat anti-rabbit secondary antibody coupled to alkaline phosphatase (Promega).

and 6). Both the cytoplasmic and membrane forms of FtsZ84, in contrast to FtsZ, failed to bind GTP on the renatured blot (Fig. 1b, lanes 6 and 7), though both forms of FtsZ84 showed up as strong immunoreactive bands (Fig. 1c, lanes 6 and 7). This indicated that GTP-binding to FtsZ is related to its ability to initiate septation. As a positive control, recombinant human p21 Ha-Ras protein¹³ bound GTP on the same blot under similar assay conditions (Fig. 1b, lane 1). Note that p21Ras did not crossreact with anti-FtsZ antibody (Fig. 1c, lane 1). As observed with FtsZ84, the products of temperature-sensitive and dominant-lethal alleles of *ras*-related yeast genes *SEC4* (ref. 14) and *YPT1* (ref. 15), respectively, fail to bind GTP *in vitro*.

The membrane-bound forms of both FtsZ and FtsZ84 showed retarded mobility on SDS gels, as revealed by GTP-blotting (Fig. 1b) and antibody staining (Fig. 1c). Such altered mobility was seen with proteins expressed either from the T7 promoter (Fig. 1c, lanes 5 and 7) or the protein produced from the

chromosomal copy of *ftsZ* (Fig. 1c, lane 3). This result indicates that both FtsZ and FtsZ84 may undergo a post-translational modification for membrane attachment.

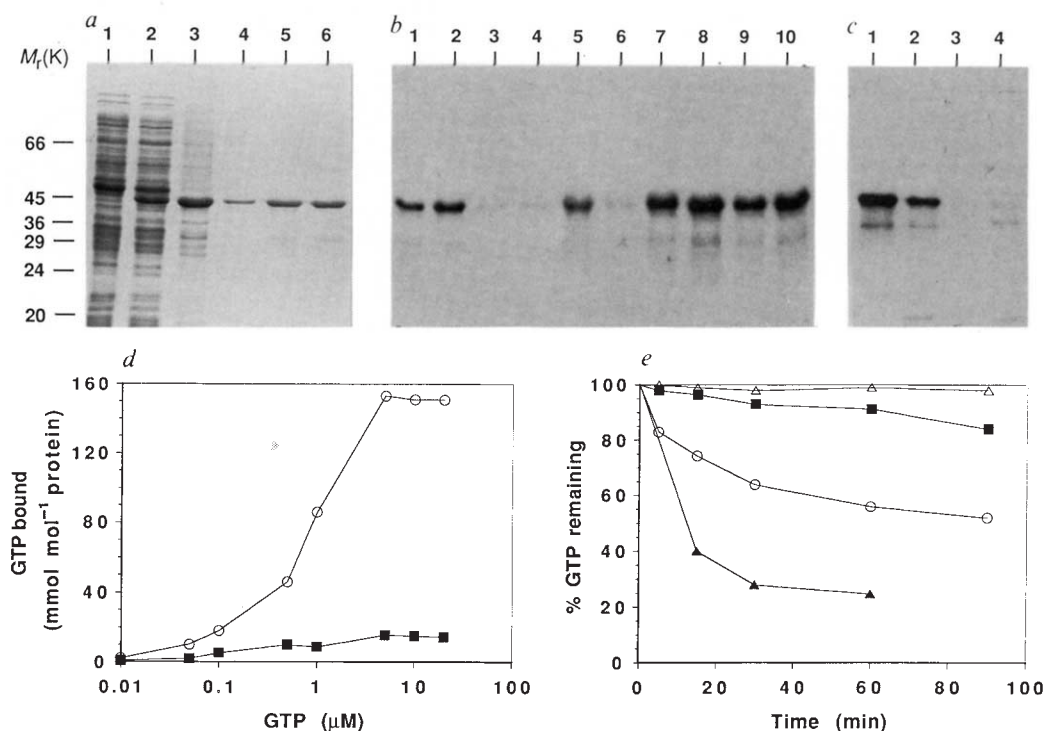
In another approach, purified soluble FtsZ protein (>95% purity, Fig. 2a) could be crosslinked using ultraviolet radiation to [α -³²P]GTP (Fig. 2b, c). Such photoaffinity labelling was stronger in the absence of Mg²⁺, probably because of an enhanced exchange of bound nucleotides¹⁶. The GTP-cross-linked band could be immunostained with anti-FtsZ antibody, confirming that the labelled protein was FtsZ. The specificity of GTP-crosslinking was tested in the presence of EDTA at 50-fold molar excess of nucleotide competitors. As shown in Fig. 2b, GTP, GTP γ S, and GDP could compete effectively, whereas dGTP, GMP, ATP, CTP and UTP failed to abolish GTP-crosslinking. This clearly demonstrated that FtsZ is a GTP/GDP-binding protein. The absence of dGTP competition (lane 5) is in marked contrast to its ability to bind to p21

FIG. 2 Purification of FtsZ

and FtsZ84 proteins and analysis of their GTP-binding and GTPase activities. a, Protein purification steps shown on a Coomassie blue-stained 12% SDS polyacrylamide gel. Lanes 1, 2, soluble lysate from IPTG-induced pET-3 control cells (15 μ g protein) and pET-3Z⁺ cells (15 μ g), respectively; lane 3, 35% (NH₄)₂SO₄ fraction of pET-3Z⁺ soluble lysate (10 μ g); lanes 4 and 5, FtsZ peak fraction from DEAE-Sephacel column, 2.5 and 5 μ g, respectively; lane 6, FtsZ84 DEAE-Sephacel peak fraction (5 μ g). b, 12% SDS-PAGE profile of UV-crosslinking of [α -³²P]GTP to FtsZ in the presence of competitors; lane 1, 5 mM MgCl₂, 0.1 mM EDTA; lane 2, 1 mM EDTA; lanes 3–10, as lane 2, with 50 μ M unlabelled nucleotide competitor; 3, GTP; 4, GTP γ S; 5, dGTP; 6, GMP; 7, GDP; 8, ATP; 9, CTP; 10, UTP. c, [α -³²P]GTP-crosslinking to purified FtsZ84; lanes 1 and 3, 2.5 mM EDTA; lanes 2, 4, and 5 mM MgCl₂, 0.1 mM EDTA; lanes 1, 2 contained purified FtsZ as control; 3, 4 contained FtsZ84. d, Filter-retention assay of [α -³²P]GTP-binding to FtsZ (○) and FtsZ84 (■). e, Time course for GTPase activity of 5 μ M FtsZ (○), 5 μ M FtsZ in the presence of 10 mM EDTA (△), 10 μ M FtsZ (▲) and 5 μ M FtsZ84 (■).

[γ -³²P]GTP hydrolysis was assayed with 5 μ M protein, whereas [α -³²P]GTP was the substrate for 10 μ M FtsZ.

METHODS. a, FtsZ and FtsZ84 proteins were overproduced at 37 or 30 °C, respectively, in cultures grown in NZ amine-yeast extract (NZY), 0.2% glucose, 100 μ g ml⁻¹ ampicillin medium (500 ml) as described in Fig. 1. The cells were collected, washed and resuspended in 10 ml of 40 mM Tris-HCl, pH 8.0, 5 mM EDTA, 140 mM NaCl, 1 mM DTT, 10% sucrose, 200 μ g ml⁻¹ PMSF, and 2 μ g ml⁻¹ Leupeptin, and treated with 400 μ g ml⁻¹ lysozyme for 60 min on ice, followed by two cycles of freeze-thawing and brief sonication. The lysate was made 10 mM with MgCl₂, clarified by centrifugation (4 °C, 60 min) at 100,000g, and the supernatant adjusted to 35% saturation with (NH₄)₂SO₄. The resulting precipitate was dissolved in 1 ml buffer A (50 mM Tris HCl, pH 7.5, 5 mM MgCl₂, 0.1 mM EDTA, 0.5 mM DTT, 10% glycerol, 200 μ g ml⁻¹ PMSF and 2 μ g ml⁻¹ Leupeptin), and dialysed. The nondialysable fraction (~10 mg protein) was loaded at 4 °C on a DEAE-Sephacel column (~15 ml bed volume) equilibrated in buffer A. The column was washed with six volumes of buffer and developed with a gradient (120 ml) of 0–1 M KCl. Both FtsZ and FtsZ84 eluted as a peak between 50 and 250 mM KCl. The



column fractions were monitored by dot blots followed by western blots, using anti-FtsZ antibody as in Fig. 1. b, c, For ultraviolet-crosslinking of GTP, 1- μ g protein samples were incubated in 30 μ l of 50 mM Tris-HCl, pH 7.8, 100 mM potassium acetate, containing MgCl₂ or EDTA as above, and 1 μ M [α -³²P]GTP (3,000 Ci mmol⁻¹) for 20 min on ice, 2.5 cm from an ultraviolet lamp. The rest of the procedure was essentially as described³⁷. d, Filter-binding reactions were done in duplicate with 0.5 μ M FtsZ or FtsZ84 in 50 μ l buffer B (50 mM Tris-HCl, pH 7.5, 50 mM KCl, 5% glycerol) containing 0.5 mM MgCl₂ and various concentrations of [α -³²P]GTP (17,600 c.p.m. pmol⁻¹). After incubation at 30 °C for 10 min, the samples were rapidly filtered without dilution through nitrocellulose discs (Millipore; 0.22 μ m) presoaked in binding buffer. The filters were washed with 2–3 vols 300 μ l cold buffer B and counted. Washing reduced the background to ~0.05% of input radioactivity. Molecular concentrations of proteins were estimated assuming *M_r*, 43K for FtsZ. e, To measure GTPase activity, proteins were incubated at 30 °C in 25 μ l buffer B supplemented with 2 mM MgCl₂ and 0.5 mM [α -³²P] or [γ -³²P]GTP (550 c.p.m. pmol⁻¹). At the times indicated, 4- μ l aliquots were added to an equal volume of 1% SDS, 20 mM EDTA, heated at 75 °C for 2 min and the nucleotides resolved by chromatography on polyethyleneimine cellulose (Merck) in 0.75 M KH₂PO₄ (pH 3.4). GTP, GDP or P_i spots were visualized with autoradiography and identified by standard markers. The radioactive spots were excised and counted to quantify hydrolysis.

TABLE 1 Putative binding domains in FtsZ

Consensus	G1 GXXGXGK ^S _T	G2 D(X) _n T	G3 DXXG	G4 NK _Q XD _{TQ}
FtsZ (<i>E. coli</i>)	¹⁰⁶ GGTGTGAA	¹²² DLGILT ¹⁵⁸ DSLIT ²¹² DVRT	¹⁸⁰ DAFG ²⁵³ DLSG	²⁸⁵ TSLD
EF-Tu (<i>E. coli</i>)	¹³ GHVDHGKT	⁵⁰ D(X) ₁₀ T	⁸⁰ DCPG	¹³⁵ NKCD
CDC42 (yeast)	¹⁰ GDGAVGKT	³² YVPT	⁵⁷ DTAG	¹¹⁵ TQID
SEC4 (yeast)	²⁷ GDSGVGKS	⁴⁸ FITT	⁷⁵ DTAG	¹³³ NKSD
Ha-Ras (human)	¹⁰ GAGGVGKS	³³ DPT	⁵⁷ DTAG	¹¹⁶ NKCD

Comparison of the putative G1–G4 GTP-binding domains in the FtsZ amino-acid sequence^{6,7} with the conserved sequence motifs in the GTPase superfamily⁸ (single-letter amino-acid code). For references to the listed sequences (except FtsZ), see ref. 8. The protein databases PIR, SwissProt, and GenPept (translated GenBank) were searched for homology to FtsZ at the National Centre for Biotechnology Information using the BLAST network service. The search did not reveal any similarity between FtsZ and the known GTP-binding proteins.

Ha-Ras¹⁰ or YPT1 (ref. 15) proteins. FtsZ84 protein, purified in an identical manner to FtsZ (Fig. 2a), did not show any significant crosslinking to [α -³²P]GTP either in the presence of Mg²⁺ or EDTA (Fig. 2c, lanes 3 and 4).

To obtain a quantitative measure of GTP-binding, we used purified proteins in a filter-binding assay (Fig. 2d). The kinetics of binding for reactions containing 1 μ M FtsZ and 5 μ M GTP revealed that maximal binding occurred within 10 min of incubation at 30 °C (not shown). As measured by ³²P-retention on filters, a maximal binding of 150 nmol GTP mol⁻¹ FtsZ was obtained in 10 min and half-maximal binding to FtsZ occurred at $\sim 8 \times 10^{-7}$ M GTP (Fig. 2d). FtsZ84 had diminished ability to bind [α -³²P]GTP even at GTP concentrations where FtsZ reached equilibrium binding (Fig. 2d). This confirmed the blotting and crosslinking results showing that FtsZ84 is defective for GTP-binding *in vitro*.

To determine whether FtsZ had GTPase activity in solution at 0.5 mM GTP concentration, thin-layer chromatography analysis was done to assess the conversion of GTP to GDP or the release of the γ -phosphate group of GTP. As shown in Fig. 2e, 5 μ M FtsZ hydrolysed 25% GTP in 15 min (2.5 nmol hydrolysed out of a total of 10 nmol GTP present in the reaction), whereas a sevenfold reduced hydrolysis (3.5%) was obtained with a similar concentration of FtsZ84 during this initial period. FtsZ-catalysed hydrolysis occurred with no detectable lag and was abolished in the presence of excess EDTA (Fig. 2e), indicating that Mg²⁺ is required. No hydrolysis occurred in the control, in the absence of FtsZ or FtsZ84.

The elongation factor Tu (EF-Tu), the most abundant protein in *E. coli* cells and a classic GTP-binding protein, has an apparent M_r similar to that of FtsZ¹⁷. To rule out the possibility of contamination of the FtsZ preparation with EF-Tu, we found that affinity-purified anti-EF-Tu antibody did not crossreact with purified FtsZ. Furthermore, purified EF-Tu failed to crossreact with anti-FtsZ antibody, and it did not show up on the GTP-blot (Fig. 1b, lanes 2 and 3).

Guanine nucleotides that were copurified with FtsZ, in the absence of any exogenously added nucleotides, had an identical retention time as GDP on a high-performance liquid chromatography (HPLC) anion-exchange column (Fig. 3). The stoichiometry of bound GDP to FtsZ was ~ 0.7 mol per mol protein. Such a predominantly GDP-bound state of FtsZ is similar to that known for EF-Tu¹⁷, yeast Ras¹⁸, and viral Ha-Ras¹⁹ proteins, and suggests that a similar FtsZ-specific GTPase cycle to that known for signal-transducing GTP-binding proteins^{8,9} occurs *in vivo*. We failed to detect any nucleotide bound to purified FtsZ84 isolated from cells grown at the permissive

temperature (30 °C). This implies a labile association of guanine nucleotides with FtsZ84 and corroborates its impaired GTP-binding ability *in vitro*.

Although the FtsZ sequence^{6,7} has certain features (G2- and G3-like motifs) in common with the canonical GTPases⁸, it deviates significantly from the consensus motifs in the putative G1 and G4 domains (Table 1). The G1 domain (residues 106–113; GGTGTGAA; single-letter code) contains Ala 112 and Ala 113 in place of the usual Lys and Ser or Thr, respectively, whereas the putative G4 region (residues 296–299; TSLD) contains Ser 297 instead of the Lys or Gln present in the consensus sequence (Table 1). The *ftsZ*84 thermosensitive mutation (Gly 105 to Ser change)⁷ and the lethal *ftsZ*3 allele (Thr 108 to Ala change)⁷ map adjacent to, or in, the putative G1 region. The mutant phenotypes indicate that this glycine-rich region is crucial for FtsZ function. There is a near-perfect match of the FtsZ residues 105–111 (GGGTGTG) with a glycine-rich region (G/AGGTGSG) present in α - and β -tubulin²⁰ as well as γ -tubulin²¹ sequences (residues 140–148). This region in tubulin chains is postulated to be the phosphate-binding loop, owing to its similarity to the nucleotide-binding folds of adenylate kinase, dehydrogenases, and flavodoxins²⁰. The temperature-sensitive *ftsZ*22 mutation is an Asp 212 to Gly substitution⁷ in a G2-like motif. On the basis of the above, we propose that FtsZ defines a new class of GTP/GDP-binding proteins. The *ftsZ*-like genes are widespread among the eubacteria²², and *E. coli ftsZ*

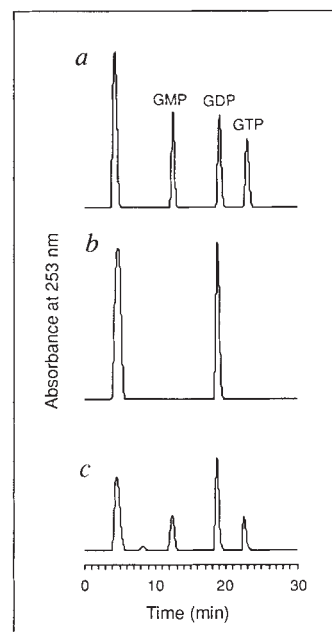


FIG. 3 HPLC analysis of the nucleotide content of purified FtsZ. a, Elution profile for a mixture of authentic GMP, GDP and GTP, 400 pmol of each in buffer A (see Fig. 2, legend). b, Elution profile of the filtrate of centrifuged, heat-denatured FtsZ. c, Elution profile of a mixture of GMP, GDP and GTP standards (each at 400 pmol) and the filtrate of centrifuged, heat-denatured FtsZ.

METHODS. HPLC was done on a Nucleogen-DEAE 500–7 column (6 mm $\times$ 125 mm; Macherey-Nagel) using Rainin software (flow rate 1 ml min⁻¹; temperature ~ 25 °C). Nucleotides were eluted with the following washes and gradients: 20 mM K₂PO₄, pH 7.0, for 3 min, 20–500 mM KPO₄ over 30 min, then 20 mM phosphate in 2 min, and finally 20 mM phosphate for 20 min, before the next sample injection. To analyse for FtsZ-bound nucleotides, the protein (200 μ g in 200 μ l of buffer A) was boiled for 3 min with 5 mM EDTA, and the denatured protein was removed by ultrafiltration using a 30K cutoff membrane (Ultrafree-MC filter unit, Millipore). The filtrate was used to obtain an ultraviolet spectrum which was characteristic of guanine and the nature of the presumed guanine nucleotide was analysed by HPLC. Retention times of guanine nucleotide standards were: 5'GMP, 12.3 min; GDP, 18.5 min; GTP, 22.3 min. The peak near the injection corresponds to components of the sample buffer such as dithiothreitol¹⁸.

homologues have been cloned from *Bacillus subtilis*²³ and *Rhizobium meliloti*²⁴. An alignment of FtsZ sequences from the three bacterial species shows that the putative G1 motif (GGTGTGAA) is perfectly conserved²⁴. We suggest that FtsZ homologues in *B. subtilis* and *R. meliloti* may also be GTP-binding proteins. If true, this would imply a common signalling mechanism for the initiation of bacterial septation.

Genes encoding Ras-related GTP-binding proteins (BUD1/RSR1, refs 25, 26; CDC42, ref. 27) and a putative GDP-GTP exchange factor (BUD5, ref. 28) are necessary for bud-site selection and assembly in yeast. Moreover, CDC3, CDC10, CDC11 and CDC12 gene products, required for cytokinesis, constitute the 10-nm filament ring in the neck connecting the mother yeast cell to its bud²⁹. Our finding that FtsZ is a GTP-binding protein and the recent evidence of its localization at the septal site as a ring-like structure³⁰ suggest that the initiation of cell division in *E. coli* and budding yeast may have common features. It is conceivable that activation of FtsZ by the exchange of bound GDP by GTP, signals the ring formation at the septal site, similar to the polymerization of ATP-bound forms of actin³¹ and *E. coli* RecA³² or the GTP-bound form of tubulin³³ into filaments. Genetic evidence implicates *ftsZ* as the target of two division-inhibitors, DNA damage-inducible Sula, and MinCD, which prevents septation at old sites^{2,3}. These inhibitors could arrest division by antagonizing GTP-mediated FtsZ activation and/or signal transduction.

As with the signalling GTPases^{8,9}, binding and hydrolysis of GTP by FtsZ during the cell cycle could be regulated by specific protein ligands, such as a guanine nucleotide exchange factor and a GTPase activating protein. Analysis of the GTPase cycle of FtsZ may reveal shared features between the primordial division machinery in *E. coli* and the eukaryotic cell division apparatus, and may offer insights into bacterial cell division. □

Received 20 May; accepted 31 July 1992.

- Donachie, W. D., Begg, K. J. & Sullivan, N. F. In *Microbial Development* (ed. Losick, R. & Shapiro, L.) 27–62 (Cold Spring Harbor Laboratory, New York, 1984).
- de Boer, P. A. J., Cook, W. R. & Rothfield, L. I. *Rev. Genet.* **24**, 249–274 (1990).
- Lutkenhaus, J. *Trends Genet.* **6**, 22–25 (1990).
- Begg, K. J. & Donachie, W. D. *J. Bact.* **163**, 615–622 (1985).
- Taschner, P. E. M., Huls, P. G., Pas, E. & Woldringh, C. L. *J. Bact.* **170**, 1533–1540 (1988).
- Yi, Q.-M. & Lutkenhaus, J. *Gene* **36**, 241–247 (1985).
- Bi, E. & Lutkenhaus, J. *J. Bact.* **172**, 5602–5609 (1990).
- Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **349**, 117–127 (1991).
- Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **348**, 125–132 (1990).
- McGrath, J. P., Capon, D. J., Goeddel, D. V. & Levinson, A. D. *Nature* **310**, 644–649 (1984).
- Lapetina, E. G. & Reep, B. R. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2261–2265 (1987).
- Connolly, T. & Gilmore, R. *Cell* **57**, 599–610 (1989).
- Farnsworth, C. L., Marshall, M. S., Gibbs, J. B., Stacey, D. W. & Feig, L. A. *Cell* **64**, 625–633 (1991).
- Goud, B., Salminen, A., Walworth, N. C. & Novick, P. J. *Cell* **53**, 753–768 (1988).
- Schmitt, H. D., Wagner, P., Pfaff, E. & Gallwitz, D. *Cell* **47**, 401–412 (1986).
- Hall, A. & Self, A. J. *J. Biol. Chem.* **261**, 10963–10965 (1986).
- Hershey, J. W. B. In *Escherichia coli & Salmonella typhimurium Cellular & Molecular Biology* (eds Neidhardt, F. C. et al.) 613–647 (American Society for Microbiology, Washington DC, 1987).
- Temeles, G. L., Gibbs, J. B., D'Alonzo, J. S., Sigal, I. S. & Scolnick, E. M. *Nature* **313**, 700–703 (1985).
- Poe, M., Scolnick, E. M. & Stein, R. B. *J. Biol. Chem.* **260**, 3906–3909 (1985).
- Kraus, E. et al. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4156–4160 (1981).
- Oakley, C. E. & Oakley, B. R. *Nature* **338**, 662–664 (1989).
- Corton, J. C., Ward, J. E. Jr & Lutkenhaus, J. *J. Bact.* **169**, 1–7 (1987).
- Beall, B., Lowe, M. & Lutkenhaus, J. *J. Bact.* **170**, 4855–4864 (1988).
- Margolin, W., Corbo, J. C. & Long, S. R. *J. Bact.* **173**, 5822–5830 (1991).
- Bender, A. & Pringle, J. R. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9976–9980 (1989).
- Chant, J. & Herskowitz, I. *Cell* **65**, 1203–1212 (1991).
- Johnson, D. & Pringle, J. R. *J. cell Biol.* **111**, 143–152 (1990).
- Chant, J., Corrado, K., Pringle, J. R. & Herskowitz, I. *Cell* **65**, 1213–1224 (1991).
- Kim, H. B., Haarer, B. K. & Pringle, J. R. *J. cell Biol.* **112**, 535–544 (1991).
- Bi, E. & Lutkenhaus, J. *Nature* **354**, 161–164 (1991).
- Janmey, P. A. et al. *Nature* **347**, 95–99 (1990).
- McEntee, K., Weinstock, G. M. & Lehman, I. R. *J. Biol. Chem.* **256**, 8835–8844 (1981).
- Snyder, J. A. & McIntosh, J. R. A. *Rev. Biochem.* **45**, 699–720 (1976).
- Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. *Meth. Enzym.* **185**, 60–89 (1990).
- Saiki, R. K. et al. *Science* **239**, 487–491 (1988).
- Ito, K., Sato, T. & Yura, T. *Cell* **11**, 551–559 (1977).
- Rosen, B. P., Weigel, U., Karkaria, C. & Gangola, P. *J. Biol. Chem.* **263**, 3067–3070 (1988).

ACKNOWLEDGEMENTS. We thank A. Wright, P. Polaczek, P. Ljungdahl, A. Antebi and L. Feig for encouragement, help and discussions; P. Ljungdahl for sharing his expertise in database search; A. Wright, L. Feig, A. L. Sonenshein, P. Polaczek, P. Ljungdahl, A. Antebi and A. N. Chatterjee for comments on the manuscript; L. Feig, J. Lutkenhaus, D. L. Miller and F. W. Studier for reagents, and A. G. Gilman for suggesting the analysis of protein-bound nucleotides. This work was supported in part by a grant from the NIH to J.T.P.

The essential bacterial cell-division protein FtsZ is a GTPase

Piet de Boer, Robin Crossley & Lawrence Rothfield

Department of Microbiology, University of Connecticut Health Center, Farmington, Connecticut 06030, USA

CYTOKINESIS defines the last stage in the division cycle, in which cell constriction leads to the formation of daughter cells. The biochemical mechanisms responsible for this process are poorly understood. In bacteria, the *ftsZ* gene product, FtsZ, is required for cell division^{1–3,17}, playing a prominent role in cytokinesis. The cellular concentration of FtsZ regulates the frequency of division⁴ and genetic studies have indicated that it is the target of several endogenous division inhibitors⁵. At the time of onset of septal invagination, the FtsZ protein is recruited from the cytoplasm to the division site, where it assembles into a ring that remains associated with the leading edge of the invaginating septum until septation is completed⁶. Here we report that FtsZ specifically binds and hydrolyses GTP. The reaction can be dissociated into a GTP-dependent activation stage that is markedly affected by the concentration of FtsZ, and a hydrolysis stage in which GTP is hydrolysed to GDP. The results indicate that GTP binding and hydrolysis are important in enabling FtsZ to support bacterial cytokinesis, either by facilitating the assembly of the FtsZ ring and/or by catalysing an essential step in the cytokinetic process itself.

The FtsZ peptides from the three bacterial species of known sequence^{7–9} share a 13 amino-acid segment that is completely conserved and that includes a 7 amino-acid stretch almost identical to the most highly conserved sequence among the α -, β - and γ -tubulins of eukaryotic cells (Fig. 1). It is proposed that this tubulin signature sequence¹⁰ is involved in the binding of GTP by tubulins^{11,12}, prompting us to investigate whether FtsZ might be a GTPase.

Evidence that FtsZ is a GTP-binding protein was obtained from experiments in which FtsZ was specifically crosslinked to [α -³²P]GTP by ultraviolet irradiation (Fig. 2). When crude *Escherichia coli* cell extracts containing elevated concentrations of endogenously generated FtsZ were subjected to the crosslinking procedure, SDS-gel electrophoresis revealed a labelled band whose position corresponded to that of FtsZ. The extent of labelling paralleled the concentration of FtsZ in the extracts (Fig. 2a, b, lanes 1–4). The ability to bind GTP was retained by the purified protein (Fig. 2a, b, lane 5) and did not require Mg²⁺ (Fig. 2c, lane 1). The nucleotide-binding reaction was specific for GTP and GDP, as shown by competition experiments with nonradioactive nucleotides (Fig. 2c, lanes 2–6) and by the inability to crosslink [α -³²P]ATP to FtsZ (data not shown).

A single amino-acid substitution within the tubulin signature sequence of FtsZ (FtsZ84; Fig. 1) leads to a failure to initiate septum formation in cells grown at elevated temperatures. As shown in Fig. 2b (lanes 6, 7), the efficiency of crosslinking of [α -³²P]GTP to the mutant FtsZ84 protein was significantly lower



FIG. 1 Tubulin signature sequence in FtsZ. The stretch of 13 amino acids that is completely conserved among FtsZ peptides is shown for the three bacterial species for which the sequence is known (*Escherichia coli*⁷, *Bacillus subtilis*⁸ and *Rhizobium meliloti*⁹). The substitution of serine for glycine in the mutant FtsZ84 protein of *E. coli*¹⁴ leads to a failure in initiating cytokinesis when mutant cells are grown at elevated temperatures¹⁵. The lower portion of the figure shows the 7 amino-acid tubulin signature sequence¹⁰ that is conserved among tubulins of higher organisms.

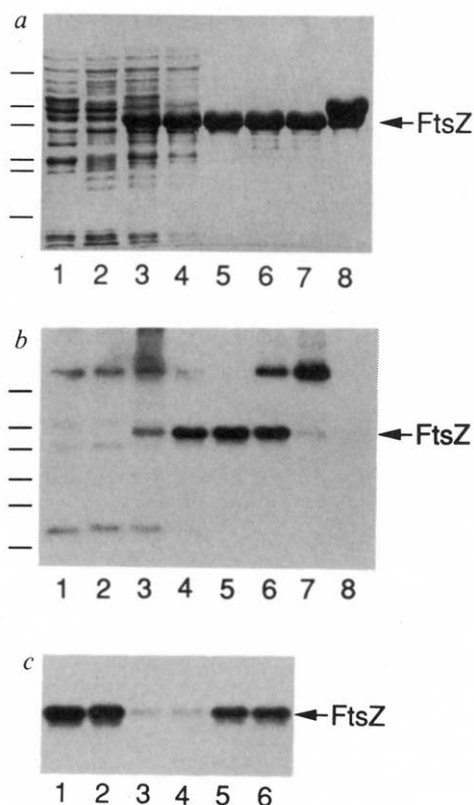


FIG. 2 Covalent crosslinking of [α - 32 P]GTP to FtsZ. *a*, Coomassie blue-stained SDS gels showing the protein profiles of the samples used in this study. Lanes 1–4, S200 extracts of strains with different levels of *ftsZ* expression: lane 1, PB115/pBSKS[pBluescriptKS vector]; lanes 2–4, PB115/pDB189[P_{lac}::ftsZ]. Cells were grown in LB medium containing 0.2% glucose to repress baseline *ftsZ* expression (lane 2), in LB medium (lane 3) or in LB medium containing 0.84 mM isopropyl- β -D-thiogalactoside (IPTG) (lanes 1, 4). Lane 5, purified FtsZ protein. Lane 6, partially purified FtsZ protein. Lane 7, partially purified FtsZ84 protein. 50 μ g protein was applied to lanes 1–4 and 30 μ g to lanes 5–7. Lane 8, 60 μ g chicken ovalbumin. The positions of protein molecular mass standards (66, 45, 36, 29, 24 and 20K, from top to bottom) are indicated to the left of the gel. *b*, The samples shown in the corresponding lanes of *a* were added to 0.5 μ M [α - 32 P]GTP (400 Ci mmol $^{-1}$), 20 mM Tris-acetate (pH 7), 100 mM potassium acetate, 1 mM EDTA, 0.5 mM dithiothreitol (DTT) and 10 mM magnesium acetate, then immediately irradiated with ultraviolet light for 30 min at 4 °C as described¹⁶ before analysis by SDS-gel electrophoresis and autoradiography. The gel lanes contained 7.1 μ g (lanes 1–4), 5.7 μ g (lane 5), 2.9 μ g (lanes 6, 7) and 10 μ g (lane 8) of protein, respectively. Lanes 6 and 7 show that a minor 75K contaminant peptide, which very efficiently forms photoadducts with [α - 32 P]GTP, is present in the partially purified preparations of both FtsZ (lane 6) and FtsZ84 (lane 7). This protein is not present in the more purified preparation of FtsZ (lane 5). *c*, Purified FtsZ (lane 5 of panel *a*) was irradiated with ultraviolet light in the presence of [α - 32 P]GTP as described in *b*. Lane 1, Mg $^{2+}$ was omitted. Lane 2, complete reaction mixture. Lanes 3–6, the following unlabelled nucleotides (50 μ M) were added to the complete reaction before crosslinking: lane 3, GTP; lane 4, GDP; lane 5, GMP; lane 6, ATP. Addition of non-radioactive dGTP, CTP or UTP gave results identical to those shown in lanes 5 and 6.

METHODS. Purified FtsZ was prepared from cells of an overproducing *E. coli* strain (PB115/pDB189) grown at 37 °C in LB medium containing 0.84 mM IPTG. Cells were broken in a French pressure cell, centrifuged for 180 min at 200,000g, the supernatant (S200) was fractionated by precipitation with ammonium sulphate, followed by column chromatography on Q-Sepharose, HA-Ultragel and Superose-12. Partially purified FtsZ and FtsZ84 proteins were prepared from the S200 fraction of PB115/pDB189[P_{lac}::ftsZ] and PB115/pDB254[P_{lac}::ftsZ84] cells grown at 30 °C by precipitation with ammonium sulphate, followed by column chromatography on Superose-12. All parameters during the preparation of partially purified FtsZ and FtsZ84 were identical. The *ftsZ84* allele in pDB254 was obtained by homologous recombination between pCX41 (*ftsZ*⁺) and the chromosome of strain AW40, *ftsZ84*, (ref. 3).

than that of the wild-type protein. Thus a single mutation affected both the biological function and the GTP-binding ability of the protein.

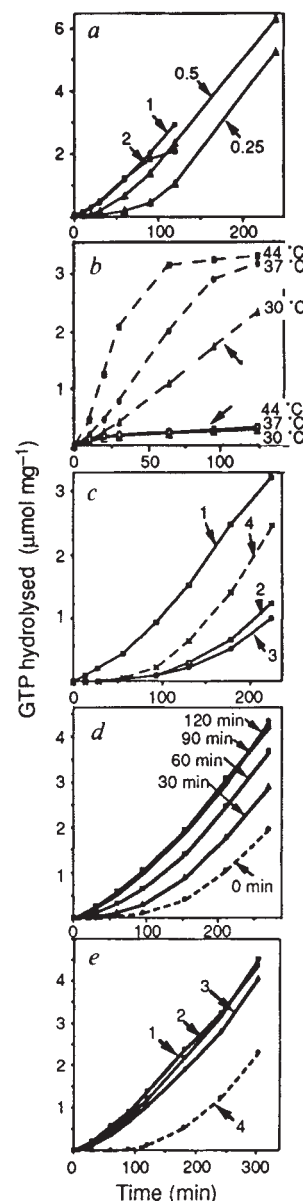
Evidence that FtsZ is a GTPase was obtained by measuring the ability of the purified protein to convert [α - 32 P]GTP to [α - 32 P]GDP (Fig. 3*a*). In contrast to the binding of GTP to FtsZ, already described, FtsZ-mediated hydrolysis of GTP required the presence of Mg $^{2+}$.

A comparison of mutant and wild-type proteins revealed that the GTPase activity of the mutant FtsZ84 protein was significantly lower than that of wild-type FtsZ (Fig. 3*b*). The mutant protein also failed to show the normal temperature-dependent increase in activity that is seen with the wild-type enzyme.

The GTPase activity of FtsZ showed an unusual dependence on protein concentration. As shown in Fig. 3*a*, a lag period occurred before the onset of significant GTP hydrolysis, the duration of which was inversely related to FtsZ concentration.

FIG. 3 GTPase activity of FtsZ. *a*, GTPase activity at different concentrations of FtsZ. Purified FtsZ (see Fig. 2*a*, lane 5) was assayed at 37 °C for GTPase activity at 0.25, 0.5, 1 and 2 mg ml $^{-1}$ protein. The reaction mixtures were equilibrated at 37 °C for 5 min before starting the reaction by the addition of Mg $^{2+}$. *b*, Mutant compared with wild-type FtsZ. Assays were done on partially purified FtsZ and FtsZ84 proteins (described in Fig. 2*a*, lanes 6 and 7) at 44 °C, 37 °C and 30 °C. The reactions were started by addition of protein (1 mg ml $^{-1}$) to the other components of the reaction mixture that had been preincubated for 10 min at the indicated temperatures. Dashed lines, FtsZ; solid lines, FtsZ84. *c*, Effect of preincubation. Purified FtsZ (0.3 mg ml $^{-1}$) was preincubated at 37 °C for 80 min in reaction mixtures from which Mg $^{2+}$ (reaction 1), GTP (reaction 2) or both Mg $^{2+}$ and GTP (reaction 3) were omitted. The reactions were then started (at time 0) by addition of the missing components. There was no detectable hydrolysis of GTP during the preincubation with [α - 32 P]GTP in the absence of Mg $^{2+}$ (reaction 1). Reaction 4 was started by addition of FtsZ to the other components of the reaction mixture that had been preincubated for 80 min. *d*, Time course of activation. Purified FtsZ was incubated with GTP before addition of Mg $^{2+}$ as described for reaction 1 in *c*. The time of preincubation varied from 0 to 120 min as indicated in the figure. *e*, Effect of GTP concentration during preincubation. Purified FtsZ was incubated with GTP for 90 min, as described for reaction 1 in *c*. The concentrations of GTP present during the preincubation period were 5 mM (reaction 1), 1 mM (reaction 2), 0.5 mM (reaction 3), or 0 mM (reaction 4). After the preincubation period the reactions were started (time 0) by addition of GTP to 5 mM, and magnesium acetate to 15 mM.

METHODS. For GTPase assays the complete reaction mixture contained 5 mM [α - 32 P]GTP (40 mCi mmol $^{-1}$), 15 mM magnesium acetate, reaction buffer (40 mM Tris-acetate (pH 7), 200 mM potassium acetate, 2 mM EDTA, 1 mM DTT and 0.5% Triton X-100) and protein as indicated. At intervals, aliquots from the reaction mixtures were subjected to thin layer chromatography and the amount of GTP converted to GDP was determined¹⁶.



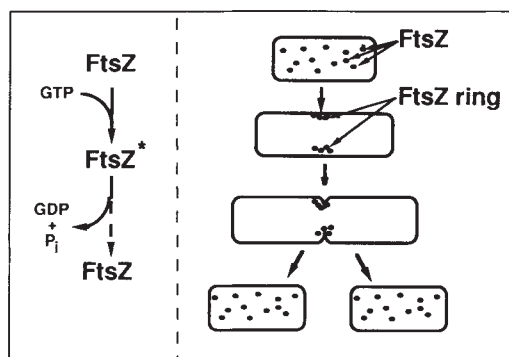


FIG. 4 *In vitro* properties and *in vivo* behaviour of FtsZ. The left panel indicates the two stages in the *in vitro* FtsZ GTPase reaction as described in the text. The right panel indicates the *in vivo* behaviour of FtsZ during the cell cycle⁶. FtsZ* indicates the activated form of FtsZ.

With longer incubation period, the specific activity of the FtsZ GTPase reached the same level at all protein concentrations as shown by the parallel slopes of the activity curves in the time-course plots. We conclude that during the lag period an activation step occurred, leading to formation of a fully functional enzyme, and that the duration of the activation step was significantly affected by protein concentration.

To determine which component of the reaction mixture was required for the activation step, purified FtsZ was preincubated with the different individual components of the mixture before the reaction was started by addition of the missing ingredients. As shown in Fig. 3c, preincubation with GTP resulted in virtual loss of the lag period whereas preincubation with the other components of the reaction mixture failed to do so. Preincubation for 90 min was required for maximal activation with 5 mM GTP (Fig. 3d). There was no measurable hydrolysis of GTP during the preincubation. These results define two stages in the GTPase reaction: an activation stage that requires binding of the nucleotide and whose duration is strongly dependent on protein concentration, and a Mg^{2+} -dependent catalytic stage in which GTP is hydrolysed to GDP.

A 10-fold decrease in GTP concentration during the preincubation, which should greatly decrease the rate of interaction of GTP with FtsZ, did not significantly affect the extent of activation, as measured by suppression of the lag period (Fig. 3e). This suggests that the binding of FtsZ to GTP is not rate-limiting in formation of the active enzyme. In contrast, as shown in Fig. 3a, small changes in protein concentration had a marked effect on the time needed to activate the enzyme, as indicated by the length of the lag.

This behaviour resembles that of self-associating systems such as the nucleotide-dependent assembly of microtubules and actin filaments¹³, and indicates that formation of the active FtsZ GTPase is dependent on interactions between FtsZ subunits, perhaps involving the assembly of FtsZ into a multimeric complex. In gel filtration studies, purified FtsZ behaved as a series of oligomers ranging in size from 6-mers to 25-mers, with a mean M_r of 400 K (10-mer) (data not shown). Whether binding of GTP alters the oligomeric state of FtsZ remains to be tested directly.

Recent immunolocalization studies⁶ defined several stages in the behaviour of FtsZ during the division cycle (Fig. 4). In the first stage, FtsZ moves from its normal cytoplasmic location to assemble at the inner surface of the cytoplasmic membrane into a ring that extends circumferentially around the cell at the site of division. This is followed by the cytokinesis stage, during

which FtsZ remains at the leading edge of the invaginating septum. When septation is completed the protein disappears from this site and is once again found in the cytoplasm.

It is likely that the GTPase activity of FtsZ plays a role in one or more of these stages. It is tempting to speculate that the concentration-dependent activation step, induced by GTP binding in the *in vitro* system, may be related to a multimerization step that leads to production of the FtsZ ring *in vivo*. The subsequent hydrolysis of GTP may play a role in the cytokinetic event itself and/or in the release of FtsZ that occurs after septation is completed. The assignment of a biochemical activity to FtsZ should permit an *in vitro* approach to studying aspects of the division process which have previously been inaccessible. □

Received 10 June; accepted 5 August 1992.

1. Dai, K. & Lutkenhaus, J. *J. Bact.* **173**, 3500–3506 (1991).
2. Pla, J., Sánchez, M., Palacios, P., Vicente, M. & Aldea, M. *Molec. Microbiol.* **5**, 1681–1686 (1991).
3. Wang, X., de Boer, P. A. J. & Rothfield, L. I. *EMBO J.* **10**, 3363–3372 (1991).
4. Ward, J. E. & Lutkenhaus, J. *Cell* **42**, 941–949 (1985).
5. de Boer, P. A. J., Cook, W. R. & Rothfield, L. I. *A. Rev. Genet.* **24**, 249–274 (1990).
6. Bi, E. & Lutkenhaus, J. *Nature* **354**, 161–164 (1991).
7. Yi, Q.-M. & Lutkenhaus, J. *Gene* **36**, 241–247 (1985).
8. Beall, B., Lowe, M. & Lutkenhaus, J. *J. Bact.* **170**, 4855–4864 (1988).
9. Margolin, W., Corbo, J. C. & Long, S. R. *J. Bact.* **173**, 5822–5830 (1991).
10. Bairoch, A. *Nucleic Acids Res.* **19**, 2241–2245 (1991).
11. Sternlicht, H., Yaffe, M. B. & Farr, G. W. *FEBS Lett.* **214**, 226–235 (1987).
12. Hesse, J., Thierauf, M. & Ponstingl, H. *J. biol. Chem.* **262**, 15472–15475 (1987).
13. Carlier, M.-F. *Int. Rev. Cytol.* **115**, 139–170 (1989).
14. Bi, E. & Lutkenhaus, J. *J. Bact.* **172**, 5602–5609 (1990).
15. Lutkenhaus, J. F., Wolf-Watz, H. & Donachie, W. C. *J. Bact.* **142**, 615–620 (1980).
16. de Boer, P. A. J., Crossley, R. E., Hand, A. R. & Rothfield, L. I. *EMBO J.* **10**, 4371–4380 (1991).
17. Beall, B. & Lutkenhaus, J. *Genes Dev.* **5**, 447–455 (1991).

ACKNOWLEDGEMENTS. This work was supported by grants from the NIH.

GUIDE TO AUTHORS

PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without interruptions of the text, contains about 1,300 words. Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and will where necessary be sent between offices by overnight courier. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, who are chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are typeset from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose, and to mention availability of these data.

Once a manuscript is accepted for publication, contributors will receive galley proofs in about 4 weeks. *Nature's* staff will edit manuscripts with a view to brevity and clarity, so contributors should check their proofs carefully. Manuscripts are generally published 2–3 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review articles survey recent developments in a field. Most are commissioned, but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items (figures and tables) and should not occupy more than five pages of *Nature*. Articles start with a heading of 50–80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarize its results and implications. Articles have fewer than 50 references and may contain a few subheadings of two or three words. **Letters** are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of non-specialist readers. Letters do not have subheadings and contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research. Some are commissioned but suggestions can be made to the commentary editor in the form of a one-page synopsis. Commentaries are normally between one and four pages of *Nature*.

News and Views articles inform non-specialist readers about new scientific advances, sometimes in the form of a conference report. Most are commissioned but proposals can be made in advance to the

News and Views editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words and 5 references.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and three copies are required, each accompanied by artwork. If photographs are included, four sets of originals are required: for line drawings, one set of originals and three good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Relevant manuscripts in press or submitted for publication elsewhere should be included with each copy of a submitted manuscript, and clearly marked as such. Revised and resubmitted manuscripts should also be clearly marked as such and labelled with their manuscript numbers.

Titles say what the paper is about with the minimum of technical terminology and in fewer than 80 characters in the case of Articles and Letters. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Abstract should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column with *Nature* can accommodate 20 amino acids or 60 base pairs. Numbering of sequences should be in the left-hand margin only, with a single space between rows.

Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome. **Colour artwork.** A charge of £500 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent the publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text. **References** are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published in or in the press are numbered and included in the reference list. All other forms of reference, including unrefereed abstracts, should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1983–85). The first and last page numbers are included; reference to books should include publisher, place and date.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used except for changed addresses.

Acknowledgements are brief; grant and contribution numbers are not allowed.

Submission. Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 117 National Press Building, Washington, DC 20045, USA. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax (VAT). □